

ПРИКЛАДНАЯ ДИСКРЕТНАЯ МАТЕМАТИКА

Научный журнал

2011

№3(13)

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



ТОМСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

**РЕДАКЦИОННАЯ КОЛЛЕГИЯ ЖУРНАЛА
«ПРИКЛАДНАЯ ДИСКРЕТНАЯ МАТЕМАТИКА»**

Агибалов Г. П., д-р техн. наук, проф. (председатель); Девянин П. Н., д-р техн. наук, проф. (зам. председателя); Парватов Н. Г., канд. физ.-мат. наук, доц. (зам. председателя); Черемушкин А. В., д-р физ.-мат. наук, чл.-корр. Академии криптографии (зам. председателя); Панкратова И. А., канд. физ.-мат. наук, доц. (отв. секретарь); Алексеев В. Б., д-р физ.-мат. наук, проф.; Бандман О. Л., д-р техн. наук, проф.; Евдокимов А. А., канд. физ.-мат. наук, проф.; Евтушенко Н. В., д-р техн. наук, проф.; Закревский А. Д., д-р техн. наук, проф., чл.-корр. НАН Беларуси; Костюк Ю. Л., д-р техн. наук, проф.; Логачев О. А., канд. физ.-мат. наук, доц.; Матросова А. Ю., д-р техн. наук, проф.; Микони С. В., д-р техн. наук, проф.; Салий В. Н., канд. физ.-мат. наук, проф.; Сафонов К. В., д-р физ.-мат. наук, проф.; Фомичев В. М., д-р физ.-мат. наук, проф.; Чеботарев А. Н., д-р техн. наук, проф.; Шоломов Л. А., д-р физ.-мат. наук, проф.

Адрес редакции: 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 надежности, интеллектуальных системах.

Периодичность выхода журнала: 4 номера в год.

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

Подписано к печати 26.08.2011.
Формат $60 \times 84\frac{1}{8}$. Усл. п. л. 13,8. Уч.-изд. л. 15,47. Тираж 300 экз.

Издательство ТГУ. 634029, Томск, ул. Никитина, 4
Отпечатано в типографии ТПУ.

СОДЕРЖАНИЕ

ТЕОРЕТИЧЕСКИЕ ОСНОВЫ ПРИКЛАДНОЙ ДИСКРЕТНОЙ МАТЕМАТИКИ

Алексейчук А. Н., Шевцов А. С. Быстрый алгоритм статистического оценивания максимальной несбалансированности билинейных аппроксимаций булевых отображений.....	5
Вороненко А. А. О сложности доказательства повторности булевых функций в бинарном базисе.....	12
Гречников Е. А. Метод комплексного умножения для построения эллиптических кривых и его оптимизации	17

МАТЕМАТИЧЕСКИЕ ОСНОВЫ КОМПЬЮТЕРНОЙ БЕЗОПАСНОСТИ

Стефанцов Д. А. Внедрение политик безопасности в программные системы обработки информации	55
--	----

ПРИКЛАДНАЯ ТЕОРИЯ ГРАФОВ

Быкова В. В. Вычислительные аспекты древовидной ширины графа	65
Ильев В. П., Навроцкая А. А. Вычислительная сложность задачи аппроксимации графами с компонентами связности ограниченного размера	80
Магомедов А. М., Магомедов Т. А. Интервальная на одной доле правильная реберная 5-раскраска двудольного графа.....	85
Монахова Э. А. Структурные и коммуникативные свойства циркулянтных сетей	92

ВЫЧИСЛИТЕЛЬНЫЕ МЕТОДЫ В ДИСКРЕТНОЙ МАТЕМАТИКЕ

Фомичев В. М. О реализации метода согласования в криптоанализе с помощью параллельных вычислений	116
---	-----

МАТЕМАТИЧЕСКИЕ ОСНОВЫ ИНТЕЛЛЕКТУАЛЬНЫХ СИСТЕМ

Киселев И. С. Аналитический метод доопределения кратных предпочтений в матрице парных сравнений	122
--	-----

СВЕДЕНИЯ ОБ АВТОРАХ	129
---------------------------	-----

АННОТАЦИИ СТАТЕЙ НА АНГЛИЙСКОМ ЯЗЫКЕ	130
--	-----

CONTENTS

THEORETICAL BACKGROUNDS OF APPLIED DISCRETE MATHEMATICS

Alekseychuk A. N., Shevtsov A. S. Fast algorithm for statistical estimation of the maximal imbalance of bilinear approximations of Boolean mappings	5
Voronenko A. A. On the complexity of proving that a Boolean function is not a binary read-once	12
Grechnikov E. A. Method for constructing elliptic curves using complex multiplication and its optimizations	17

MATHEMATICAL BACKGROUNDS OF COMPUTER SECURITY

Stefantsov D. A. Implementation of security policies in programming information processing systems	55
---	----

APPLIED GRAPH THEORY

Bykova V. V. Computational aspects of graph treewidth	65
Il'ev V. P., Navrocka A. A. Computational complexity of the problem of approximation by graphs with connected components of bounded size	80
Magomedov A. M., Magomedov T. A. Regular edge 5-coloring of bipatite graph that is an interval on one part	85
Monakhova E. A. Structural and communicative properties of circulant networks	92

COMPUTATIONAL METHODS IN DISCRETE MATHEMATICS

Fomichev V. M. On implementation of the meet-in-the-middle attack by means of parallel computations	116
--	-----

MATHEMATICAL BACKGROUNDS OF INTELLIGENT SYSTEMS

Kiselev I. S. Analytical method for making definite multiple preferences in the matrix of incomplete pairwise comparisons	122
--	-----

BRIEF INFORMATION ABOUT THE AUTHORS	129
---	-----

PAPER ABSTRACTS	130
-----------------------	-----

ТЕОРЕТИЧЕСКИЕ ОСНОВЫ ПРИКЛАДНОЙ ДИСКРЕТНОЙ МАТЕМАТИКИ

DOI 10.17223/20710410/13/1

УДК 631.391:519.2

БЫСТРЫЙ АЛГОРИТМ СТАТИСТИЧЕСКОГО ОЦЕНИВАНИЯ МАКСИМАЛЬНОЙ НЕСБАЛАНСИРОВАННОСТИ БИЛИНЕЙНЫХ АППРОКСИМАЦИЙ БУЛЕВЫХ ОТОБРАЖЕНИЙ

А. Н. Алексейчук, А. С. Шевцов

Институт специальной связи и защиты информации Национального технического университета Украины «Киевский политехнический институт», г. Киев, Украина

E-mail: alex-crypto@mail.ru, ashef@mail.ru

Предложен вероятностный алгоритм, позволяющий оценивать сверху максимальную несбалансированность (в заданном классе) билинейных аппроксимаций булевых отображений n переменных за время, линейно зависящее от n .

Ключевые слова: *блочный шифр, билинейный криптоанализ, булево отображение, билинейная аппроксимация, вероятностный алгоритм.*

Введение

При исследовании стойкости блочных шифров относительно билинейного метода криптоанализа требуется вычислять или оценивать максимальные значения несбалансированности билинейных аппроксимаций булевых отображений, реализуемых узлами замены, раундовой функцией или шифром в целом [1–3]. Естественные алгоритмы решения этой задачи, основанные на «прямом» вычислении искомых параметров, имеют экспоненциальную трудоемкость и становятся практически неприменимыми уже при умеренных значениях числа переменных данного отображения.

Начиная с публикации О. Гольдрайха и Л. Левина [4], известен ряд вероятностных алгоритмов [5–10], позволяющих формировать список «высоковероятных» линейных аппроксимаций произвольной булевой функции n переменных за полиномиальное от n время. Применение таких алгоритмов к линейным комбинациям координатных функций булева отображения позволяет существенно уменьшить сложность нахождения его наиболее вероятных линейных аппроксимаций (за счет некоторого снижения достоверности результата, что обусловлено вероятностным характером применяемых алгоритмов). Отметим, в частности, работу [10], где с использованием одного из таких алгоритмов получено около 80 (близких по качеству к наилучшим известным) линейных аппроксимаций шифра DES с 8 раундами шифрования.

Следует подчеркнуть, что алгоритмы, изложенные в [4–10], предназначены именно для построения линейных аппроксимаций, несбалансированности которых ограничены снизу определенным значением. Вместе с тем в ряде задач криптографии, например при обосновании стойкости блочных шифров или их элементов относительно линейных атак, требуется находить лишь нетривиальные верхние оценки максимальной несбалансированности линейных аппроксимаций. Сказанное относится и к более широкому классу билинейных аппроксимаций, для построения или оценки несбалансированности которых, по-видимому, не предлагались ранее полиномиальные алгоритмы.

В настоящей работе описан вероятностный алгоритм, позволяющий оценивать сверху максимальную несбалансированность (в заданном широком классе, см. ниже формулу (3)) билинейных аппроксимаций булевых отображений n переменных за время, линейно зависящее от n . Предложенный алгоритм базируется на развитии идеи, лежащей в основе усовершенствованного алгоритма Левина [5, 6], и может рассматриваться как обобщение одного из этапов последнего на случай билинейных аппроксимаций булевых отображений.

1. Постановка задачи и основные результаты

Обозначим V_n пространство двоичных векторов длины n , $F_{m \times n}$ — множество матриц размера $m \times n$ над полем $F = \text{GF}(2)$. Для любых $x = (x_1, \dots, x_n)$, $y = (y_1, \dots, y_n) \in V_n$ положим $xy = x_1y_1 \oplus \dots \oplus x_ny_n$, $x \oplus y = (x_1 \oplus y_1, \dots, x_n \oplus y_n)$. Строки произвольной матрицы $U \in F_{m \times n}$ обозначим U_1, \dots, U_m .

Пусть $g : V_n \rightarrow V_n$ — булево отображение, заданное с помощью оракула (некоторого алгоритма, позволяющего вычислять значение $g(x)$ по произвольному входному значению $x \in V_n$; см., например, [6]); $*$ — бинарная операция на множестве V_n . Назовем *билинейной аппроксимацией* (между входами и выходами) отображения g произвольную функцию вида

$$x \mapsto xAg(x) \oplus \alpha x \oplus \beta g(x), \quad x \in V_n, \quad (1)$$

где $A \in F_{n \times n}$, $\alpha \in V_n$, $\beta \in V_n \setminus \{0\}$. Число

$$l_*^{(g)}(A, \alpha, \beta) = 2^{-n} \sum_{k \in V_n} \left(2^{-n} \sum_{x \in V_n} (-1)^{xAg(x*k) \oplus \alpha x \oplus \beta g(x*k)} \right)^2 \quad (2)$$

назовем (средней) *несбалансированностью* указанной аппроксимации (относительно операции $*$). Далее будем отождествлять функцию (1) с упорядоченным набором (A, α, β) .

Для любого подпространства L векторного пространства V_n обозначим

$$l_*^{(g)}(L, \beta) = \max\{l_*^{(g)}(A, \alpha, \beta) : A_1, \dots, A_n \in L, \alpha \in V_n\} \quad (3)$$

максимальную несбалансированность билинейных аппроксимаций (A, α, β) отображения g по всем векторам $\alpha \in V_n$ и $n \times n$ -матрицам A , строки которых принадлежат подпространству L . Требуется построить вероятностный алгоритм, вычисляющий для заданных $\varepsilon, \delta \in (0, 1)$ статистическую верхнюю оценку параметра (3), то есть такое случайное значение $\theta_{L, \beta} \in (0, 1)$, для которого

$$\mathbf{P}\{l_*^{(g)}(L, \beta) \leq \theta_{L, \beta} + \varepsilon\} \geq 1 - \delta. \quad (4)$$

Введем ряд дополнительных обозначений.

Для любого натурального t обозначим B_t совокупность непустых подмножеств множества $\{1, 2, \dots, t\}$. Пусть $X_1, \dots, X_t, Y_1, \dots, Y_t$ и $K^{(S)}$ ($S \in B_t$) — независимые в совокупности случайные векторы с равномерным распределением на множестве V_n . Положим $X_S = \bigoplus_{i \in S} X_i$, $Y_S = \bigoplus_{i \in S} Y_i$,

$$g_{1,S} = g(X_S * K^{(S)}), \quad g_{2,S} = g(Y_S * K^{(S)}), \quad S \in B_t. \quad (5)$$

Для любых $A \in F_{n \times n}$, $\alpha \in V_n$, $\beta \in V_n \setminus \{0\}$ зададим случайную величину

$$\xi(A, \alpha, \beta) = (2^t - 1)^{-1} \sum_{S \in B_t} (-1)^{(X_S A \oplus \beta)g_{1,S} \oplus (Y_S A \oplus \beta)g_{2,S} \oplus \alpha(X_S \oplus Y_S)}. \quad (6)$$

Заметим, что на основании равенств (2), (5), (6)

$$\mathbf{E} \xi(A, \alpha, \beta) = l_*^{(g)}(A, \alpha, \beta). \quad (7)$$

Кроме того, для любых различных множеств $S, S' \in B_t$ случайные векторы $(X_S, Y_S, K^{(S)})$ и $(X_{S'}, Y_{S'}, K^{(S')})$ независимы. Следовательно,

$$\begin{aligned} \mathbf{D} \xi(A, \alpha, \beta) &= (2^t - 1)^{-2} \sum_{S \in B_t} \mathbf{D} \left((-1)^{(X_S A \oplus \beta)g_{1,S} \oplus (Y_S A \oplus \beta)g_{2,S} \oplus \alpha(X_S \oplus Y_S)} \right) \leq \\ &\leq (2^t - 1)^{-2} (2^t - 1) \mathbf{E} (1) = (2^t - 1)^{-1}. \end{aligned} \quad (8)$$

Наконец, для любых $U, V \in F_{t \times n}$, $\beta \in V_n \setminus \{0\}$ положим

$$\eta_{(U, V, \beta)}^a = (2^t - 1)^{-1} \sum_{S \in B_t} (-1)^{(U_S \oplus \beta)g_{1,S} \oplus (V_S \oplus \beta)g_{2,S} \oplus a_S}, \quad a = (a_1, \dots, a_t) \in V_t, \quad (9)$$

где $U_S = \bigoplus_{i \in S} U_i$; $V_S = \bigoplus_{i \in S} V_i$; $a_S = \bigoplus_{i \in S} a_i$; $S \in B_t$. Отметим, что вектор, составленный из значений (9), является (с точностью до сомножителя $(2^t - 1)^{-1}$) произведением матрицы Адамара $H_t = ((-1)^{xy})_{x, y \in V_t}$ на вектор с координатами

$$g_{(U, V, \beta)}^S = \begin{cases} (-1)^{(U_S \oplus \beta)g_{1,S} \oplus (V_S \oplus \beta)g_{2,S}}, & \text{если } S \in B_t, \\ 0, & \text{если } S = \emptyset. \end{cases} \quad (10)$$

Утверждение 1. Пусть

$$\theta_{L, \beta} = \max \{ \eta_{(U, V, \beta)}^a : U_1, \dots, U_t, V_1, \dots, V_t \in L, a \in V_t \}. \quad (11)$$

Тогда для любых $\varepsilon, \delta \in (0, 1)$, удовлетворяющих условию

$$\delta^{-1} \varepsilon^{-2} \leq 2^t - 1, \quad (12)$$

справедливо неравенство (4).

Доказательство. Обозначим A^* и α^* соответственно матрицу A и вектор α , для которых достигается максимум в правой части равенства (3): $l_*^{(g)}(L, \beta) = l_*^{(g)}(A^*, \alpha^*, \beta)$. Положим $U_i^* = X_i A^*$, $V_i^* = Y_i A^*$, $a_i^* = \alpha^*(X_i \oplus Y_i)$, $i = 1, \dots, t$; $a^* = (a_1^*, \dots, a_t^*)$; обозначим U^* и V^* случайные матрицы, составленные из вектор-строк U_1^*, \dots, U_t^* и V_1^*, \dots, V_t^* соответственно.

На основании формул (6), (9) справедливо равенство $\xi(A^*, \alpha^*, \beta) = \eta_{(U^*, V^*, \beta)}^{a^*}$. При этом в силу (11) событие $\{l_*^{(g)}(L, \beta) > \theta_{L, \beta} + \varepsilon\}$ влечет событие $\{\eta_{(U^*, V^*, \beta)}^{a^*} < l_*^{(g)}(A^*, \alpha^*, \beta) - \varepsilon\}$. Отсюда, используя соотношения (7), (8) и неравенство Чебышева, получим

$$\begin{aligned} &\mathbf{P} \{ l_*^{(g)}(L, \beta) > \theta_{L, \beta} + \varepsilon \} \leq \mathbf{P} \{ \eta_{(U^*, V^*, \beta)}^{a^*} < l_*^{(g)}(A^*, \alpha^*, \beta) - \varepsilon \} = \\ &= \mathbf{P} \{ \xi(A^*, \alpha^*, \beta) < \mathbf{E} \xi(A^*, \alpha^*, \beta) - \varepsilon \} \leq \mathbf{P} \{ |\xi(A^*, \alpha^*, \beta) - \mathbf{E} \xi(A^*, \alpha^*, \beta)| > \varepsilon \} \leq \\ &\leq \varepsilon^{-2} \mathbf{D} \xi(A^*, \alpha^*, \beta) \leq \varepsilon^{-2} (2^t - 1)^{-1} \leq \delta, \end{aligned}$$

где последнее неравенство вытекает из формулы (12).

Итак, при выполнении условия (12) справедливо неравенство (4), что и требовалось доказать. ■

Полученное утверждение позволяет предложить следующий *вероятностный алгоритм вычисления верхних границ параметра* (3) по указанным выше исходным данным $g, \beta, L, \varepsilon, \delta$.

1. Положить

$$t = \lceil \log(1 + \delta^{-1}\varepsilon^{-2}) \rceil, \quad (13)$$

сгенерировать независимые в совокупности случайные векторы $X_i, Y_i, K^{(S)}$ ($i = 1, \dots, t, S \in B_t$) с равномерным распределением на множестве V_n , вычислить значения (5).

2. Для каждой пары матриц $U, V \in F_{t \times n}$, таких, что $U_i, V_i \in L$ ($i = 1, \dots, t$):

- вычислить значения (10);
- вычислить значения (9), применяя к вектору с координатами (10) алгоритм быстрого преобразования Адамара;
- положить $\theta_{U,V,\beta} = \max\{\eta_{(U,V,\beta)}^a : a \in V_t\}$.

3. Положить $\theta_{L,\beta} = \max\{\theta_{U,V,\beta} : U_i, V_i \in L (i = 1, \dots, t)\}$.

Обозначим $t^*(n)$ временную сложность операции $*$, т. е. максимальное число двоичных операций, выполняемых при вычислении значений $x * y$ для любых $x, y \in V_n$.

Утверждение 2. Пусть $\dim L = r$, где $r \equiv \text{const}$ (не зависит от n, ε, δ). Тогда временная сложность описанного алгоритма составляет

$$T_{\varepsilon,\delta}(n, r) = O((\varepsilon^{-2}\delta^{-1})^{2r+1}(n \log(\varepsilon^{-2}\delta^{-1}) + \log^2(\varepsilon^{-2}\delta^{-1}) + t^*(n))) \quad (14)$$

двоичных операций и $O(\varepsilon^{-2}\delta^{-1})$ обращений к оракулу g . При этом объем памяти, необходимой для выполнения алгоритма, равен

$$M_{\varepsilon,\delta}(n) = O(\varepsilon^{-2}\delta^{-1}(n + \log(\varepsilon^{-2}\delta^{-1}))) \text{ бит.} \quad (15)$$

Доказательство. На шаге 1 для нахождения векторов X_S, Y_S ($S \in B_t$) достаточно выполнить $O(2^t nt)$ сложений по модулю 2. Следовательно, вычисление значений (5) можно осуществить за $T_1 = O(2^t nt + 2(2^t - 1)t^*(n)) = O(2^t(nt + t^*(n)))$ двоичных операций и $O(2^t)$ обращений к оракулу g .

На шаге 2 для каждой пары матриц U, V нахождение значений (10) потребует выполнения $O(2^t nt)$ двоичных операций, а вычисление значений (9) с помощью алгоритма быстрого преобразования Адамара (без учета деления на $2^t - 1$) — $O(2^t t)$ операций сложения или вычитания целых чисел (см., например, [11], следствие 5.34). Поскольку разрядность указанных чисел не превосходит t , то двоичная временная сложность нахождения максимального значения (9) при фиксированных U, V и β не превосходит $O(2^t t(n + t))$. Наконец, так как число пар (U, V) , удовлетворяющих условию $U_i, V_i \in L$ ($i = 1, \dots, t$), равно 2^{2rt} , то суммарная временная сложность второго и третьего шагов алгоритма составляет $T_2 = O(2^{t(2r+1)}t(n + t))$ двоичных операций. Складывая выражения T_1 и T_2 , с учетом формулы (13) и условия $r \equiv \text{const}$ получим равенство (14).

Для оценки емкостной сложности алгоритма заметим, что объем памяти, необходимой для хранения чисел (5), составляет $O(2^t n)$ бит. Далее, для нахождения значения $\theta_{L,\beta}$ достаточно хранить текущие значения матриц U и V , соответствующие им

векторы (9), (10), а также ранее вычисленное значение $\theta_{\tilde{U}, \tilde{V}, \beta}$, соответствующее матрицам \tilde{U} и \tilde{V} , выбранным на предыдущем шаге вычислений. Суммарный объем необходимой для этого памяти не превосходит $O(nt + 2^t t)$ бит, откуда следует справедливость формулы (15). ■

Отметим, что в практически значимом случае, когда $t^*(n) = O(n)$ и число t вида (13) меньше n , оценки (14), (15) упрощаются и принимают следующий вид:

$$T_{\varepsilon, \delta}(n, r) = O(n(\varepsilon^{-2}\delta^{-1})^{2r+1} \log(\varepsilon^{-2}\delta^{-1})), M_{\varepsilon, \delta}(n) = O(n\varepsilon^{-2}\delta^{-1}).$$

При $L = \{0\}$, $* = \oplus$ предложенный алгоритм позволяет оценивать свеху (с точностью ε и достоверностью $1 - \delta$) максимум квадратов нормированных коэффициентов Уолша — Адамара булевой функции $f(x) = \beta g(x)$, $x \in V_n$ за $O(n\varepsilon^{-2}\delta^{-1} \log(\varepsilon^{-2}\delta^{-1}))$ двоичных операций. В этом случае предложенный алгоритм по существу совпадает с первым этапом усовершенствованного алгоритма Левина [5, 6]. Отметим, что последний алгоритм формирует случайный список, содержащий с вероятностью не менее $1 - \delta$ каждую аффинную функцию, находящуюся от функции f на расстоянии не более $2^{n-1}(1 - \varepsilon)$, со сложностью $O(n\varepsilon^{-2}\delta^{-1} \log n \log(\varepsilon^{-2}\delta^{-1}))$ операций над n -разрядными целыми числами.

2. Результаты моделирования алгоритма

Описанный алгоритм был применен к исследованию отображений $g : V_{32} \rightarrow V_{32}$, построенных по схеме блока подстановки алгоритма шифрования ГОСТ 28147-89 [12]. Вычислительные эксперименты проводились для различных наборов узлов замены $s_i : V_4 \rightarrow V_4$, $i = 0, \dots, 7$, задающих отображение g , векторов $\beta \in V_{32}$ и подпространств L размерности 0 или 1. В качестве типового примера, иллюстрирующего полученные результаты, приведем оценки параметра (3), полученные для подстановки $g = (s_0, \dots, s_7)$, операции $*$ сложения по модулю 2^{32} на множестве V_{32} , вектора

$$\beta = (1\ 0\ 0\ 0\ | 0\ 0\ 0\ 0\ | 0\ 0\ 0\ 0\ | 0\ 0\ 0\ 0\ | 0\ 0\ 0\ 0\ | 0\ 0\ 0\ 0\ | 0\ 0\ 0\ 0\ | 0\ 0\ 0\ 0) \quad (16)$$

и подпространства L , порожденного вектором

$$z = (0\ 0\ 1\ 0\ | 0\ 0\ 0\ 0\ | 0\ 0\ 0\ 0\ | 0\ 0\ 0\ 0\ | 0\ 0\ 0\ 0\ | 0\ 0\ 0\ 0\ | 0\ 0\ 0\ 0\ | 0\ 0\ 0\ 0).$$

Узлы замены s_0, \dots, s_7 в выражении подстановки g определяются по табл. 1. Они характеризуются наименьшими значениями параметров

$$d(s) = \max \{2^{-4} |\{x \in V_4 : s(x \oplus a) \oplus s(x) = b\}| : a, b \in V_4 \setminus \{0\}\},$$

$$l(s) = \max \left\{ \left(2^{-4} \sum_{x \in V_4} (-1)^{ax \oplus bs(x)} \right)^2 : a, b \in V_4 \setminus \{0\} \right\}$$

среди всех подстановок $s : V_4 \rightarrow V_4$ ($d(s_i) = l(s_i) = 0,25$ для любого $i = 0, \dots, 7$) и рекомендуются в [13] для применения в алгоритме шифрования ГОСТ 28147-89.

В табл. 2 приведены численные оценки параметра (3). Отметим, что на основании следствий 1 и 3 работы [3] точное значение этого параметра в рассматриваемом случае может быть найдено по формуле

$$l(s_0, \beta_0) = \max_{(A, \alpha)} \left\{ 2^{-4} \sum_{k \in V_4} \left(2^{-4} \sum_{x \in V_4} (-1)^{(xA s_0(x+k) \oplus \alpha x \oplus \beta_0 s_0(x+k))} \right)^2 \right\},$$

Таблица 1

Набор «экстремальных» узлов замены [13]

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
s_0	0	1	B	D	9	E	6	7	C	5	8	3	F	2	4	A
s_1	0	1	2	4	3	5	8	A	7	9	6	D	B	E	C	F
s_2	0	1	B	2	8	6	F	3	E	A	4	9	D	5	7	C
s_3	0	1	B	2	8	3	F	6	E	A	4	9	D	5	7	C
s_4	0	4	B	2	8	6	A	1	E	F	3	9	D	5	7	C
s_5	0	4	B	2	8	3	F	1	E	A	6	9	D	5	7	C
s_6	0	B	F	9	1	5	6	8	3	A	4	C	E	D	7	2
s_7	0	7	A	E	9	1	D	8	C	2	B	F	3	5	4	6

где $\beta_0 = (1, 0, 0, 0)$ — подвектор вектора (16), соответствующий подстановке s_0 ; $x + k$ — сумма по модулю 2^4 двоичных целых чисел, соответствующих векторам $x, k \in V_4$; максимум берется по всем векторам $\alpha \in V_4$ и матрицам $A \in F_{4 \times 4}$, строки которых принадлежат множеству $\{(0, 0, 0, 0), (0, 0, 1, 0)\}$. Таким образом, точное значение параметра (3) — $l(s_0, \beta_0) = 0,312500$, что отличается от его верхних оценок в среднем на 0,06 (см. последнюю колонку табл. 2).

Таблица 2

Результаты выполнения алгоритма

($\varepsilon = 0,2$, $\delta = 0,1$, $t = 8$)

№ эксперимента	$\theta_{L,\beta}$	$\theta_{L,\beta} + \varepsilon$
1	0,184314	0,384314
2	0,254902	0,454902
3	0,160784	0,360784
4	0,160784	0,360784
5	0,137255	0,337255
6	0,192157	0,392157
7	0,184314	0,384314
8	0,152941	0,352941
9	0,137255	0,337255
10	0,168627	0,368627

Для повышения точности оценок следует уменьшить значение ε (увеличить значение t) при применении алгоритма, что, очевидно, приведет к повышению его трудоемкости. При $t = 8$ время работы компьютерной программы для ЭВМ Intel Pentium Dual-Core T4300 (2,1 ГГц, 3 Гбайт RAM) с использованием пакета прикладных программ Maple 13 составляет около 26 ч.

В целом, предложенный алгоритм представляется достаточно перспективным для криптографических приложений, прежде всего, для анализа и обоснования стойкости блочных шифров относительно билинейных атак.

ЛИТЕРАТУРА

1. Courtois N. T. Feistel schemes and bi-linear cryptanalysis // Advances in Cryptology — CRYPTO'04. Springer Verlag, 2004. P. 23–40.
2. Алексейчук А. Н., Шевцов А. С. Показатели и оценки стойкости блочных шифров относительно статистических атак первого порядка // Реєстрація, зберігання і обробка даних. 2006. Т. 8. № 4. С. 53–63.

3. *Алексейчук А. Н., Шевцов А. С.* Верхние оценки несбалансированности билинейных аппроксимаций раундовых функций блочных шифров // Кибернетика и системный анализ. 2010. № 4. С. 42–51.
4. *Goldreich O. and Levin L. A.* A hard core predicate for all one-way functions // Proc. of the 21th ACM Sympos. of Theory of Computing. NY, USA: ACM, 1989. P. 25–32.
5. *Levin L. A.* Randomness and non-determinism // J. Symbolic Logic. 1993. V. 58. No. 3. P. 1102–1103.
6. *Bshouty N., Jackson J., and Tamon C.* More efficient PAC-learning of DNF with membership queries under the uniform distribution // Proc. 12th Annual Conf. on Comput. Learning Theory. NY, USA: ACM, 1999. P. 286–295.
7. *Goldreich O., Rubinfeld R., and Sudan M.* Learning polynomials with queries: the highly noisy case // SIAM J. Discrete Math. 2000. V. 13. No. 4. P. 535–570. Extended version: <http://people.csail.mit.edu/madhu/papers.html>.
8. *Kabatiansky G. and Tavernier C.* List decoding of Reed-Muller codes // Proc. Ninth Int. Workshop on Algebraic and Comb. Coding Theory. Kranevo, Bulgaria, 2004. P. 230–235.
9. *Trevisan L.* Some applications of coding theory in computational complexity // <http://eprint.arXiv:cs./0409044v1>. 24 Sept., 2004.
10. *Fourquet R., Loidreau P., and Tavernier C.* Finding good linear approximations of block ciphers and its application to cryptanalysis of reduced round DES // Proc. of the WCC 2009: ced.tavernier.free.fr/publications.
11. *Логачев О. А., Сальников А. А., Яценко В. В.* Булевы функции в теории кодирования и криптологии. М.: МЦНМО, 2004. 470 с.
12. ГОСТ 28147-89. Системы обработки информации. Защита криптографическая. Алгоритм криптографического преобразования. М.: Госстандарт СССР, 1989.
13. *Ростовцев А. Г., Маховенко Е. Б.* Введение в теорию итерированных шифров. СПб.: НПО «Мир и Семья», 2003. 302 с.

**О СЛОЖНОСТИ ДОКАЗАТЕЛЬСТВА ПОВТОРНОСТИ
БУЛЕВЫХ ФУНКЦИЙ В БИНАРНОМ БАЗИСЕ¹**

А. А. Вороненко

Московский государственный университет им. М. В. Ломоносова, г. Москва, Россия

E-mail: dm6@cs.msu.ru

Показано, что для доказательства повторности булевой функции в базисе всех функций двух переменных в худшем случае требуется линейное по числу переменных функции количество наборов.

Ключевые слова: *бесповторная булева функция, сложность доказательства.*

Булева функция $f(x_1, \dots, x_n)$, представляемая (не представляемая) бесповторной формулой в некотором базисе, называется *бесповторной (повторной)* в этом базисе. Множество всех функций двух переменных называется *бинарным базисом*.

В настоящей работе доказываются две теоремы.

Теорема 1. Существует повторная в бинарном базисе функция n переменных, доказательство повторности которой требует предъявления не менее $n + 3$ наборов.

Теорема 2. Для любой повторной в бинарном базисе функции n переменных можно предъявить $4n$ наборов, доказывающих ее повторность.

Их следствием является

Теорема 3. Доказательство повторности булевой функции в базисе всех функций двух переменных в худшем случае требует линейного относительно числа переменных функции количества наборов.

Преобразованиями обобщенной однотипности называются замена переменной или самой функции на ее отрицание и перестановка переменных. Функции, получаемые друг из друга конечным числом преобразований обобщенной однотипности, называются *обобщенно однотипными*.

Функциями Стеценко называются функции следующих пяти семейств:

$$\begin{aligned} f_d^{(s)} &= x_1 \wedge (x_2 \vee x_3 \wedge \dots \wedge x_s) \vee x_2 \wedge \bar{x}_3 \wedge \dots \wedge \bar{x}_s, & s \geq 3, \\ f_t^{(s)} &= x_1 \wedge x_2 \wedge \dots \wedge x_s \vee \bar{x}_1 \wedge \bar{x}_2 \wedge \dots \wedge \bar{x}_s, & s \geq 2, \\ f_m^{(s)} &= x_1 \wedge (x_2 \vee \dots \vee x_s) \vee x_2 \wedge \dots \wedge x_s, & s \geq 3, \\ f_4 &= x_1 \wedge (x_2 \vee x_3) \vee x_3 \wedge x_4, \\ f_5 &= x_1 \wedge (x_3 \wedge x_4 \vee x_5) \vee x_2 \wedge (x_3 \vee x_4 \wedge x_5). \end{aligned}$$

Теорема Стеценко [1]. Из любой повторной в базисе $\{\wedge, \vee, \neg\}$ функции можно подстановкой констант получить подфункцию, однотипную с одной из функций Стеценко.

¹Работа выполнена при поддержке гранта Президента РФ МД-757.2011.9.

Функциями Перязева называются следующие четыре функции:

$$\begin{aligned} p_1 &= x_1 \wedge x_2 \wedge x_3 \oplus \bar{x}_2 \wedge \bar{x}_3, \\ p_2 &= x_1 \wedge (x_2 \vee x_3) \oplus x_2 \wedge x_3, \\ p_3 &= x_1 \wedge x_2 \wedge x_3 \wedge x_4 \oplus (x_1 \oplus \bar{x}_2) \wedge \bar{x}_3 \wedge \bar{x}_4, \\ p_4 &= x_1 \wedge (x_2 \vee x_3 \wedge x_4) \oplus (x_3 \vee x_2 \wedge x_4). \end{aligned}$$

Теорема Перязева [2]. Из любой повторной в бинарном базисе функции можно подстановкой констант получить подфункцию, однотипную с одной из функций Стеценко (для $f_t^{(s)}$ выполняется $s \geq 3$), либо одну из функций Перязева.

Рассмотрим повторную в бинарном базисе функцию $(x_1 \vee \dots \vee x_n) \wedge (\bar{x}_1 \vee \dots \vee \bar{x}_n)$, где $n \geq 3$. Оценим снизу количество наборов в доказательстве ее повторности. Для доказательства повторности этой функции требуется предъявить оба ее нуля, иначе предъявляемая частичная функция доопределима либо до $x_1 \vee \dots \vee x_n$, либо до $\bar{x}_1 \vee \dots \vee \bar{x}_n$.

Назовем 0–1-матрицу *разнозначной*, если все ее строки различны и она не имеет ни нулевых, ни единичных строк. Будем говорить, что функция φ моделирует разнозначную матрицу, если $\varphi(x) = 1$ для всех строк x матрицы и $\varphi(\mathbf{0}) = \varphi(\mathbf{1}) = 0$.

Лемма 1. Каковы бы ни были три (четыре) ненулевых и неединичных набора значений трех (четырех) переменных, найдется неповторная функция, равная на них единице и нулю на единичном и нулевом наборах.

Доказательство. В силу возможности замены всех переменных их отрицаниями будем считать, что средний вес наборов не превосходит половины числа столбцов матрицы, составленной из этих наборов. В силу симметричности функции $(x_1 \vee \dots \vee x_n) \wedge (\bar{x}_1 \vee \dots \vee \bar{x}_n)$ на каждом слое булева куба, начиная с первого, будем выбирать минимальные в лексикографическом порядке наборы. Для трех переменных

возникает три матрицы $\begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}$, $\begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix}$, $\begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix}$. Первые две из них моделирует функция $x_1 \oplus (x_2 \vee x_3)$, последнюю — $x_2 \oplus x_3$.

Заметим, что часть матриц с четырьмя столбцами моделируется функцией $x_1 \oplus x_2 \oplus x_3 \oplus x_4$ либо одной из функций вида $(x_{i_1} \oplus x_{i_2}) \vee (x_{i_3} \oplus x_{i_4})$. Чтобы этого не произошло, матрица должна иметь набор веса два из каждой из трех пар противоположных наборов. Возможны четыре случая:

$$\begin{pmatrix} 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{pmatrix}, \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 \end{pmatrix}.$$

Все эти матрицы моделирует функция $(\bar{x}_1 \vee \bar{x}_2) \wedge (x_3 \vee x_4)$. ■

Для ненулевых и неединичных наборов длины не менее пяти введем следующую классификацию: наборы *первого рода* — имеющие ровно один ноль или ровно одну единицу; наборы *второго рода* — имеющие ровно два нуля или ровно две единицы (эти значения — 0 и 1 соответственно — назовём *особыми*); наборы *третьего рода* — остальные.

Лемма 2. Пусть дана произвольная разнозначная 0–1-матрица размера $m \times n$, где $m \leq n$ и $n \geq 5$. Тогда либо эта матрица моделируется некоторой бесповторной функцией, либо в ней есть два столбца, удалением которых вместе со строками, имеющими в них разные значения, и удалением дублирующихся строк можно вновь получить разнозначную матрицу с числом строк, не превосходящим числа столбцов.

Доказательство. Пусть противоположных строк второго рода нет. Рассмотрим граф с n вершинами и ребрами между теми из них, для которых существует строка второго рода, имеющая особые значения в соответствующих позициях (граф *второго рода*). Если в графе второго рода есть две неизолированные вершины, не соединенные ребром, то выбор соответствующих столбцов приводит к удалению не менее двух строк. В противной ситуации рассмотрим два случая.

1. Граф второго рода пуст. Если $m < n$, то можно взять любые два столбца так, чтобы они различались хотя бы в одной строке. Если $m = n$ и есть хотя бы две строки первого рода, то можно взять два столбца, в которых одна из них принимает уникальное значение. Наконец, при $m = n$ и наличии не более одной строки первого рода общее количество вариантов выбора пар столбцов не меньше, чем $3(n-1)(n-3) + n - 1$, что больше, чем $\binom{n}{2}$. Поэтому какая-то пара встретится не менее двух раз.

2. Граф второго рода — объединение клики размера $l \geq 2$ и изолированных вершин. Пусть в клику входят первые l вершин. Если имеется хотя бы одна строка первого или третьего рода, то, взяв два столбца: один из первых l , а другой — из оставшихся, которые отличаются в этой строке, получим утверждение леммы. Если строк первого и третьего родов нет, то при $l \geq 5$ матрицу моделирует функция

$$(x_1 \oplus x_{l+1}) \vee (x_2 \oplus x_{l+2}) \vee \cdots \vee (x_l \oplus x_{2l}).$$

При $l = 4$ с точностью до симметрий возможны два вида матриц:

$$\begin{pmatrix} 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 & 1 & 1 \\ 1 & 0 & 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 & 1 & 1 \end{pmatrix} \quad \text{и} \quad \begin{pmatrix} 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 & 1 & 1 \\ 1 & 0 & 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 1 \end{pmatrix}.$$

Данные матрицы моделируются функциями $x_6 \wedge (\overline{x_1} \vee \overline{x_2} \vee \overline{x_3} \vee \overline{x_4})$ и $(x_1 \oplus x_2) \vee (x_3 \oplus x_5)$ соответственно. ■

Для доказательства теоремы 1 достаточно воспользоваться леммами 1 и 2 и тем, что если функция φ моделирует матрицу, полученную из исходной по лемме 2 удалением i -го и j -го столбцов, то функция $(x_i \oplus x_j) \vee \varphi$ моделирует исходную матрицу.

Перейдем к доказательству верхней оценки — теоремы 2.

Дерево называется *каноническим*, если оно удовлетворяет следующим условиям:

- 1) Листья помечены попарно различными переменными или отрицаниями переменных.
- 2) Внутренние вершины помечены одним из следующих символов: \wedge , \vee , \oplus , $\overline{\oplus}$ (вершины любой степени захода, которым соответствует конъюнкция, дизъюнкция, сумма по модулю 2 или ее отрицание).

- 3) Смежные вершины не могут быть помечены одинаковыми символами из множества $\{\wedge, \vee, \oplus, \overline{\oplus}\}$, а также символами \oplus и $\overline{\oplus}$ одновременно.
- 4) Корень расположен снизу. Вершина, смежная сверху с вершиной, помеченной \oplus или $\overline{\oplus}$, не может быть помечена ни отрицанием переменной, ни конъюнкцией.

В работе [3] доказывается следующий факт.

Утверждение 1. Любая неповторная функция (зависящая хотя бы от одной переменной) в бинарном базисе представима единственным каноническим деревом.

Переменную x_i назовем *особой* для функции $f(x_1, \dots, x_n)$, если для любой другой переменной x_j подстановкой констант на места оставшихся переменных можно либо получить как линейную, так и нелинейную функции, существенно зависящие от x_i и x_j , либо получить две нелинейных функции переменных x_i и x_j , одна из которых становится константой при подстановке $x_i = 0$, другая — при подстановке $x_i = 1$.

Из утверждения 1 вытекает

Лемма 3. Бесповторная в бинарном базисе функция не имеет особых переменных.

Лемма 4. Переменная x_1 — особая для функций семейства $f_t^{(s)}$. Для доказательства этого достаточно предъявить $4s$ наборов.

Доказательство. Подстановки $x_2 = \dots = x_{j-1} = x_{j+1} = \dots = x_s = 0$ дают всевозможные подфункции вида $\overline{x_1} \wedge \overline{x_j}$, а подстановки $x_2 = \dots = x_{j-1} = x_{j+1} = \dots = x_s = 1$ — вида $x_1 \wedge x_j$. Общее число различных наборов равно $4s$. ■

Лемма 5. Переменная x_2 — особая для функций семейства $f_m^{(s)}$. Для доказательства этого достаточно предъявить $4s$ наборов.

Доказательство. Подстановки $x_1 = 1, x_3 = \dots = x_{j-1} = x_{j+1} = \dots = x_s = 0$ дают всевозможные подфункции вида $x_2 \vee x_j$ для $j \geq 3$, а подстановки $x_1 = 0, x_3 = \dots = x_{j-1} = x_{j+1} = \dots = x_s = 1$ — всевозможные подфункции вида $x_2 \wedge x_j$ для $j \geq 3$. Подстановки $x_3 = \dots = x_s = 0$ и $x_3 = \dots = x_s = 1$ дают подфункции $x_2 \wedge x_1$ и $x_2 \vee x_1$ соответственно. Общее число различных наборов равно $4s$. ■

В каноническом дереве назовем *эквивалентными* переменные или отрицания переменных, если помеченные ими вершины смежны с одной внутренней вершиной. Если вершина, помеченная x_j или $\overline{x_j}$, лежит в корневом поддереве с корнем — вершиной, смежной с вершиной, помеченной x_i или $\overline{x_i}$, то скажем, что переменная x_i не слабее переменной x_j .

Лемма 6. Для доказательства повторности функции семейства $f_d^{(s)}$ достаточно предъявить $4s$ наборов.

Доказательство. Подстановки $x_1 = 1, x_2 = 0, x_4 = \dots = x_{j-1} = x_{j+1} = \dots = x_s = 1$ дают всевозможные подфункции вида $x_3 \wedge x_j$ для $j \geq 4$, а подстановки $x_1 = 0, x_2 = 1, x_4 = \dots = x_{j-1} = x_{j+1} = \dots = x_s = 0$ — всевозможные подфункции вида $\overline{x_3} \wedge \overline{x_j}$ для $j \geq 4$. При этом подстановка $x_2 = 0, x_4 = \dots = x_s = 1$ дает остаточную функцию $x_3 \wedge x_1$, подстановка $x_1 = x_4 = \dots = x_s = 1$ — остаточную функцию $x_3 \vee x_2$, а подстановка $x_1 = 0, x_4 = \dots = x_s = 1$ — остаточную функцию $\overline{x_3} \wedge x_2$. В силу значений функции при этих подстановках переменная x_3 не может быть не слабее переменных x_4, \dots, x_s и не может быть эквивалентна переменной x_2 . Остаются неопровергнутыми два варианта для неповторной реализации.

1. Переменная x_3 эквивалентна x_1 . Тогда для гипотетической неповторной функции можно сделать замену $u = x_3 \wedge x_1$, что противоречит соотношению $f(1101\dots 1) \neq f(01\dots 1)$.

2. Переменная x_3 не слабее эквивалентных друг другу переменных x_1 и x_2 . Подстановка $x_3 = 0, x_4 = \dots = x_s = 1$ дает остаточную функцию $x_1 \wedge x_2$. Тогда для гипотетической неповторной функции можно сделать замену $u = x_1 \wedge x_2$. Противоречивым окажется соотношение $f(101\dots 1) \neq f(01\dots 1)$.

Общее число различных наборов не превосходит $4s$. ■

Лемма 7. Для доказательства повторности функции Стеценко f_5 достаточно предъявить 14 наборов.

Доказательство. Подстановка $x_2 = 0, x_4 = 1, x_5 = 0$ дает подфункцию $x_1 \wedge x_3$, подстановка $x_2 = 1, x_4 = 0, x_5 = 1$ — подфункцию $x_1 \vee x_3$. Таким образом, между x_1 и x_3 и нелинейной связкой в гипотетической неповторной формуле должны быть линейные связки, как и после нелинейной связки, т. е. формула имеет вид

$$((x_1 \oplus x_i^{\sigma_i}) \vee (x_3 \oplus x_j^{\sigma_j})) \oplus x_k^{\sigma_k},$$

где $\{i, j, k\} = \{2, 4, 5\}$. При этом остаточные функции переменных x_i, x_k и x_j, x_k будут линейными. Подстановка $x_1 = 1, x_2 = 0, x_3 = 1$ дает подфункцию $x_4 \vee x_5$, а подстановка $x_1 = 0, x_3 = 0, x_5 = 1$ — подфункцию $x_2 \wedge x_4$. Имеем противоречие. Получаемые при подстановках четверки имеют общие наборы. Общее количество различных наборов равно 14. ■

Доказательство теоремы 2. Выделим по теореме Перязева подфункцию $s \leq n$ переменных. При $s \leq 4$ теорема вытекает из неравенства $2^s \leq 4s$, при больших n — из лемм 4, 5, 6 и 7.

ЛИТЕРАТУРА

1. Стеценко В. А. О предплохих базисах в P_2 // Матем. вопросы кибернетики. Вып. 4. М.: Физматлит, 1992. С. 139–177.
2. Перязев Н. А. Слабовторные булевы функции в бинарном базисе // Дискретная математика и информатика. Вып. 4. Иркутск: Изд-во Иркут. ун-та, 1998. 12 с.
3. Вороненко А. А. О проверяющих тестах для неповторных функций // Матем. вопросы кибернетики. Вып. 11. М.: Физматлит, 2002. С. 163–176.

**МЕТОД КОМПЛЕКСНОГО УМНОЖЕНИЯ ДЛЯ ПОСТРОЕНИЯ
ЭЛЛИПТИЧЕСКИХ КРИВЫХ И ЕГО ОПТИМИЗАЦИИ¹**

Е. А. Гречников

*Московский государственный университет им. М. В. Ломоносова, г. Москва, Россия***E-mail:** grechnik@mccme.ru

Для построения эллиптических кривых над конечными полями с предписанными требованиями к их порядку используется метод комплексного умножения. В этом методе на этапе, требующем больше всего времени, вычисляется некоторый многочлен с целыми коэффициентами. В работе доказаны необходимые теоретические результаты и подробно описано, каким образом в методе комплексного умножения можно использовать делитель этого многочлена с коэффициентами в некотором расширении поля рациональных чисел.

Ключевые слова: эллиптические кривые, конечные поля, метод комплексного умножения, совместные приближения.

Введение

В последнее время эллиптические кривые играют важную роль в разнообразных приложениях, в числе которых можно назвать криптосистемы с открытым ключом [1, 2], алгоритмы разложения на множители [3] и проверки простоты [4] целых чисел. Для этих приложений необходимы эллиптические кривые над конечными полями, порядок которых удовлетворяет определённым ограничениям. Так, для криптографических применений желательно, чтобы порядок кривой был простым числом или, по крайней мере, имел по возможности больший простой делитель.

Один из методов построения эллиптических кривых с заданными условиями на порядок кривой заключается в следующем. Будем случайным образом генерировать коэффициенты уравнения, задающего кривую. Для каждого сгенерированного уравнения вычислим порядок соответствующей кривой и проверим, удовлетворяет ли он нужным условиям; если нет, перейдём к следующей кривой. Распределение значений порядка кривой при таком подходе оказывается приблизительно равномерным, строгое утверждение для случая простого поля характеристики, большей 3, можно найти в [3]. К сожалению, сложность алгоритма подсчёта порядка кривой растёт хоть и полиномиально, но достаточно быстро.

Теория комплексного умножения предоставляет другой, более практичный метод создания кривой с заданными ограничениями на порядок. Здесь сначала подбирается порядок кривой, удовлетворяющий нужным ограничениям, после чего строится кривая с заданным порядком. В п. 1 описываются детали метода и некоторые известные его оптимизации.

Предлагается новый способ оптимизации вычислений. Для его изложения понадобятся теоретические результаты, доказанные в п. 2 и 3. В п. 4 описан предлагаемый подход, в п. 5, 6 и 7 — его детали.

¹Работа поддержана грантом РФФИ № 11-01-12098.

1. Метод комплексного умножения

1.1. Теоретические основы

Будем везде считать, что $D \in \mathbb{Z}$ такое, что

$$D < 0 \text{ и либо } D \equiv 0 \pmod{4}, \text{ либо } D \equiv 1 \pmod{4}. \quad (1)$$

Рассмотрим поле $K = \mathbb{Q}(\sqrt{D})$; его дискриминант обозначим через d . Тогда $d < 0$ и

$$d \equiv 1 \pmod{4} \text{ и } d \text{ свободно от квадратов,} \quad (2)$$

— либо

$$d \equiv 0 \pmod{4}, \quad \frac{d}{4} \text{ свободно от квадратов,} \quad \frac{d}{4} \not\equiv 1 \pmod{4}. \quad (3)$$

Кроме того, $D = f^2d$, где $f \in \mathbb{N}$. Обозначим через $\mathcal{O} = \mathbb{Z} \left[\frac{d + \sqrt{d}}{2} \right]$ кольцо целых поля K и через $\mathcal{O}_D = \mathbb{Z} \left[\frac{D + \sqrt{D}}{2} \right]$ порядок в \mathcal{O} кондуктора f . Если M — произвольное числовое поле, то через \mathcal{O}_M будем обозначать кольцо целых этого поля; так, $\mathcal{O}_K = \mathcal{O}$.

Дробным идеалом порядка \mathcal{O}_D будем называть подмножество K , являющееся конечно порождённым \mathcal{O}_D -модулем и содержащее ненулевой элемент; целым идеалом или просто идеалом \mathcal{O}_D — дробный идеал, содержащийся в \mathcal{O}_D (что совпадает с обычным определением идеала кольца с исключением нулевого идеала). Собственным (дробным) идеалом порядка \mathcal{O}_D будем называть (дробный) идеал \mathfrak{a} , такой, что $\{\beta \in K : \beta\mathfrak{a} \subset \mathfrak{a}\} = \mathcal{O}_D$. Все собственные дробные идеалы порядка \mathcal{O}_D образуют абелеву группу относительно операции умножения идеалов [5, §7], которую будем обозначать $I(\mathcal{O}_D)$. Легко видеть, что главные дробные идеалы, т. е. множества вида $\alpha\mathcal{O}_D$ с $\alpha \in K^*$, образуют подгруппу в $I(\mathcal{O}_D)$, которую будем обозначать $P(\mathcal{O}_D)$. Будем называть эквивалентными идеалы, отличающиеся умножением на главный дробный идеал; запись $\mathfrak{a} \sim \mathfrak{b}$ будет обозначать эквивалентность дробных идеалов \mathfrak{a} и \mathfrak{b} . Будем для краткости называть класс эквивалентных собственных дробных идеалов просто классом идеалов. Поскольку $P(\mathcal{O}_D)$ — подгруппа, то классы идеалов образуют фактор-группу $\mathcal{H}_D = I(\mathcal{O}_D)/P(\mathcal{O}_D)$. Она называется группой классов идеалов и является конечной абелевой группой [5, §7]. Поскольку \mathcal{O}_D и K инвариантны относительно комплексного сопряжения, то комплексное сопряжение дробного идеала как множества само является дробным идеалом; операция комплексного сопряжения идеалов индуцирует корректно определённую операцию на \mathcal{H}_D .

Квадратичной формой назовём выражение вида $Ax^2 + Bxy + Cy^2$, где $A, B, C \in \mathbb{Z}$. Будем также обозначать такие выражения тройками (A, B, C) . Дискриминантом квадратичной формы назовём величину $B^2 - 4AC$, две формы назовём эквивалентными, если одна переходит в другую заменой переменных (x, y) с целочисленной матрицей определителя 1. Назовём квадратичную форму положительно определённой, если её дискриминант отрицателен и $A > 0$, и примитивной, если $\gcd(A, B, C) = 1$. Далее будем рассматривать только примитивные положительно определённые квадратичные формы дискриминанта D , для краткости не оговаривая этого явно. Корнем формы будем называть (единственный) корень τ уравнения $A\tau^2 + B\tau + C = 0$, лежащий в верхней комплексной полуплоскости $\mathbb{H} = \{z \in \mathbb{C} : \text{Im } z > 0\}$, т. е. $\tau = \frac{-B + \sqrt{D}}{2A} \in K \cap \mathbb{H}$. Приведённой формой назовём такую форму, что $|B| \leq A \leq C$, и если $B < 0$, то $|B| < A < C$. Каждая форма эквивалентна ровно одной приведённой форме [5, Theorem 2.8].

Элементы группы \mathcal{H}_D находятся во взаимно-однозначном соответствии с приведёнными формами, которое будем обозначать буквой \mathfrak{h} , а именно [5, Theorem 7.7]: форме $\xi = (A, B, C)$ с корнем τ можно сопоставить класс $\mathfrak{h}(\xi) = \mathfrak{h}(A, B, C)$ идеалов \mathcal{O}_D , содержащий $\langle 1, \tau \rangle_{\mathbb{Z}}$ (который является собственным дробным идеалом \mathcal{O}_D), причём эквивалентным формам будет сопоставлен один и тот же класс.

Все приведённые формы легко перечислить: нетрудно видеть, что для такой формы справедливы неравенства $|B| \leq A \leq \sqrt{|D|/3}$ и при фиксированных A, B существует не более одного варианта для C . Это делает приведённые формы удобным способом задания элементов \mathcal{H}_D .

Наряду с классическим определением модулярного инварианта j как функции на верхней полуплоскости \mathbb{H} [6, §46] можно также определить j -инвариант решётки в \mathbb{C} [5, §10], не меняющийся при умножении решётки на любое ненулевое комплексное число, так, что при $\tau \in \mathbb{H}$ выполнено равенство $j(\tau) = j(\langle 1, \tau \rangle_{\mathbb{Z}})$. Если \mathfrak{a} — собственный дробный идеал \mathcal{O}_D , то он является решёткой в \mathbb{C} . При этом умножение идеала \mathfrak{a} на главный идеал приводит к умножению решётки на элемент из K^* ; следовательно, значение $j(\mathfrak{a})$ зависит только от класса идеала \mathfrak{a} . С вычислительной точки зрения если дробный идеал \mathfrak{a} принадлежит классу, заданному формой $Ax^2 + Bxy + Cy^2$ с корнем τ , то есть $\mathfrak{a} \sim \langle 1, \tau \rangle_{\mathbb{Z}}$, то $j(\mathfrak{a}) = j(\tau)$.

Для любой эллиптической кривой при $n \in \mathbb{Z}$ определим отображение $[n]$, переводящее точку P в точку nP . В частности, отображение $[1]$ — тождественное. Изогенией двух эллиптических кривых будем называть морфизм (в смысле алгебраической геометрии), переводящий бесконечно удалённую точку одной кривой в бесконечно удалённую точку другой, эндоморфизмом эллиптической кривой будем называть изогению в себя. При любом $n \in \mathbb{Z}$ отображение $[n]$ является эндоморфизмом эллиптической кривой [7, Example III.4.1] и коммутирует с любым другим эндоморфизмом (поскольку любая изогения эллиптических кривых является гомоморфизмом групп точек в силу [7, Theorem III.4.8]); кольцо эндоморфизмов эллиптической кривой является \mathbb{Z} -модулем относительно действия $n\varphi = [n] \circ \varphi$, где $n \in \mathbb{Z}$, φ — эндоморфизм. Эндоморфизмы $\{[n] : n \in \mathbb{Z}\}$ образуют кольцо, изоморфное \mathbb{Z} [7, Proposition III.4.2].

Кольцо эндоморфизмов эллиптической кривой над \mathbb{C} либо совпадает с $\{[n] : n \in \mathbb{Z}\}$, либо изоморфно порядку в некотором мнимоквадратичном поле [7, Corollary III.9.4 и Exercise 3.18b]. В последнем случае говорят, что кривая имеет комплексное умножение на этот порядок. Есть ровно $|\mathcal{H}_D|$ неизоморфных эллиптических кривых с умножением на \mathcal{O}_D [7, Proposition C.11.1], эти кривые характеризуются тем, что их j -инвариант совпадает со значением модулярного инварианта $j(\mathfrak{a})$, вычисленного на представителях классов идеалов \mathcal{O}_D . Эти значения называются сингулярными значениями. Любое такое значение генерирует над K одно и то же поле $L = L(D) = K(j(\mathfrak{a}))$, называемое полем классов кольца \mathcal{O}_D (*ring class field*) [5, Theorem 11.1]. Группа Галуа расширения L/K изоморфна группе \mathcal{H}_D [5, §9]. Канонический изоморфизм, который будем обозначать Ω , сопоставляет классу идеалов с представителем \mathfrak{b} автоморфизм, переводящий $j(\mathfrak{a})$ в $j(\mathfrak{a}\mathfrak{b}^{-1})$ [5, Corollary 11.37]. Комплексное сопряжение действует следующим образом: $j(\mathfrak{a}) = j(\bar{\mathfrak{a}})$ по определению j [5, §10], $\mathfrak{a}\bar{\mathfrak{a}} \sim \mathcal{O}_D$ [5, (7.6)], следовательно, $j(\bar{\mathfrak{a}}) = j(\mathfrak{a}^{-1})$.

Рассмотрим многочлен $H_D[j](x) = \prod_{i=1}^h (x - j(\mathfrak{a}_i))$, где $h = |\mathcal{H}_D|$ и \mathfrak{a}_i — представители всех классов идеалов. Его коэффициенты лежат в L , инвариантны относительно действия $\text{Gal}(L/K)$ и комплексного сопряжения, то есть лежат в \mathbb{Q} . Более того, $j(\mathfrak{a}_i)$ — целые алгебраические числа [5, Theorem 11.1], так что $H_D[j](x) \in \mathbb{Z}[x]$.

Пусть p простое, n натуральное, $q = p^n$, эллиптическая кривая E определена над конечным полем \mathbb{F}_q . Под точками и эндоморфизмами кривой E при отсутствии явного указания основного поля будем понимать $\overline{\mathbb{F}}_q$ -точки и эндоморфизмы над $\overline{\mathbb{F}}_q$. Под порядком кривой будем понимать число \mathbb{F}_q -точек. Кольцо эндоморфизмов $\text{End}(E)$ изоморфно либо порядку в некотором мнимоквадратичном поле, либо порядку в некоторой алгебре кватернионов над \mathbb{Q} [7, Corollary III.9.4 и Theorem V.3.1]. В последнем случае кривая E называется суперсингулярной, и такие кривые нас интересовать не будут. В первом случае $\text{End}(E) \cong \mathcal{O}_D$ для некоторого D , удовлетворяющего (1); $\text{End}(E) = \langle [1], \alpha \rangle_{\mathbb{Z}}$, где α — некоторый эндоморфизм кривой E . Оказывается [8, Теорема 13.14], что кривая вместе с эндоморфизмом α может быть «поднята» в \mathbb{C} в следующем смысле: существует числовое поле L' , кривая E' , определённая над ним, эндоморфизм α' кривой E' , идеал $\mathfrak{B}' \subset \mathcal{O}_{L'}$, лежащий над p (то есть $\mathfrak{B}' \cap \mathbb{Z} = p\mathbb{Z}$), и редукция E' по модулю \mathfrak{B}' , изоморфная E , причём при этой редукции α' переходит в α . Поскольку $\alpha \notin \{[n] : n \in \mathbb{Z}\}$, то и $\alpha' \notin \{[n] : n \in \mathbb{Z}\}$, так что $\text{End}(E') \not\cong \mathbb{Z}$ и E' обладает комплексным умножением на некоторый порядок в некотором мнимоквадратичном поле. Согласно свойствам редукции [8, Теорема 13.12], она индуцирует изоморфизм $\text{End}(E')$ с подкольцом в $\text{End}(E)$. Поскольку редукция α' есть α , то $\text{End}(E') \cong \text{End}(E) \cong \mathcal{O}_D$.

В $\text{End}(E)$ есть эндоморфизм Фробениуса $Fr : (x, y) \mapsto (x^q, y^q)$ и дуальный к нему эндоморфизм \widehat{Fr} , причём $Fr \circ \widehat{Fr} = [q]$ [7, Theorem III.6.2 и Proposition 2.11] и $|\text{E}(E_q)| = |\text{Ker}([1] - Fr)| = ([1] - Fr) \circ ([1] - \widehat{Fr})$ (первое равенство следует из того, что Fr оставляет неподвижными точки \mathbb{F}_q и только их, второе — [7, Theorem III.4.10, Corollary III.5.5, Theorem III.6.2]). Обозначим элемент $\mathcal{O}_D \cong \text{End}(E)$, соответствующий Fr , через π , а элемент, соответствующий \widehat{Fr} , через $\bar{\pi}$; тогда $\pi\bar{\pi} = q$ и $(1-\pi)(1-\bar{\pi}) = |E(\mathbb{F}_q)|$. В частности, при $\pi \notin \mathbb{R}$ из этих равенств легко видеть, что $\bar{\pi}$ — действительно комплексное сопряжение к π ; если $\pi \in \mathbb{R}$, то $\pi \in \mathbb{R} \cap \mathcal{O}_D = \mathbb{Z}$, следовательно, $Fr = [\pi]$, а в таком случае $\widehat{Fr} = Fr$ [7, Theorem III.6.2], так что и в этом случае $\bar{\pi}$ совпадает с комплексным сопряжением к π .

Поскольку $\pi \in \mathcal{O}_D$, то существуют $u, v \in \mathbb{Z}$, такие, что $\pi = (u + v\sqrt{D})/2$. Тогда $\bar{\pi} = (u - v\sqrt{D})/2$ и $q = \pi\bar{\pi} = (u^2 - Dv^2)/4$, т. е. $4q = u^2 + |D|v^2$. Порядок кривой E при этом равен $1 - \pi - \bar{\pi} + \pi\bar{\pi} = q + 1 - u$; в силу [7, Exercise 5.10] из несуперсингулярности следует, что $(q, u) = 1$.

Подытожим вышесказанное: если E — несуперсингулярная эллиптическая кривая над \mathbb{F}_q , то найдутся число D , числовое поле L' и кривая E' над ним, такие, что

- E' обладает комплексным умножением на \mathcal{O}_D ;
- существует редукция E' , изоморфная E ;
- порядок E равен $q + 1 - u$, где $u \in \mathbb{Z}$ такое, что для некоторого $v \in \mathbb{Z}$ выполнено равенство $4q = u^2 + |D|v^2$.

1.2. Базовый алгоритм

Для построения таких кривых E реализуем следующую схему, подробности которой будут описаны позже:

- 1) Выберем числа $q = p^n$, p простое, и $\hat{u}, \hat{v}, D \in \mathbb{Z}$ так, чтобы D удовлетворяло (1) и выполнялись следующие условия:

$$4q = \hat{u}^2 + |D|\hat{v}^2; \quad (4)$$

$$\text{gcd}(\hat{u}, q) = 1, \quad (5)$$

а также чтобы порядок поля q и порядок будущей кривой $q + 1 - \hat{u}$ удовлетворяли условиям, нужным для конкретных приложений.

- 2) Вычислим многочлен $H_D[j](x)$.
- 3) Редуцируем многочлен $H_D[j](x)$ по модулю p ; получим некоторый многочлен над $\mathbb{F}_p \subset \mathbb{F}_q$, который, как будет показано, разлагается на линейные множители над \mathbb{F}_q . Вычислим какой-нибудь из его корней и построим эллиптическую кривую E'' над \mathbb{F}_q с j -инвариантом, равным вычисленному корню.
- 4) Кривая E'' обладает комплексным умножением на \mathcal{O}_D . Изоморфизм не меняет кольца комплексного умножения, но может изменить число \mathbb{F}_q -точек. В дальнейшем будет показано, как построить кривую, изоморфную E'' , с числом точек $q + 1 - \hat{u}$.

Под $\left(\frac{a}{p}\right)$, где $a \in \mathbb{Z}$, p — простое, будем понимать символ Кронекера, который при нечётном p равен символу Лежандра, а при $p = 2$ определён только в случае $a \equiv 0, 1 \pmod{4}$ и равен

$$\left(\frac{a}{2}\right) = \begin{cases} 1, & \text{если } a \equiv 1 \pmod{8}, \\ -1, & \text{если } a \equiv 5 \pmod{8}, \\ 0, & \text{если } a \equiv 0 \pmod{4}. \end{cases}$$

При $b \in \mathbb{N}$ символ Кронекера $\left(\frac{a}{b}\right)$ определим по мультипликативности.

Условия (4) и (5) накладывают довольно сильные ограничения на q и p . В частности, справедлива следующая лемма.

Лемма 1. Пусть $d < 0$ удовлетворяет условию (2) или (3), $D = df^2$. Пусть для $q = p^n$ и некоторых целых u, v выполнены условия $4q = u^2 + |D|v^2$, $\gcd(q, u) = 1$. Тогда справедливы следующие утверждения:

1)

$$\left(\frac{D}{p}\right) = \left(\frac{d}{p}\right) = 1; \quad (6)$$

2) $p \nmid f$;

3) $p\mathcal{O} = \mathfrak{p}\bar{\mathfrak{p}}$, где $\mathfrak{p} \neq \bar{\mathfrak{p}}$ — простые идеалы \mathcal{O} ;

4) $\frac{u + v\sqrt{D}}{2} \in \mathcal{O}_D$;

5) $\frac{u + v\sqrt{D}}{2}\mathcal{O} = \mathfrak{p}^n$ или $\frac{u + v\sqrt{D}}{2}\mathcal{O} = \bar{\mathfrak{p}}^n$.

Доказательство. Из равенства $4q = u^2 + |D|v^2$ и условия $p \nmid u$ легко видеть, что $p \nmid D$ и $p \nmid \hat{v}$; кроме того, редукция равенства по модулю p даёт $u^2 - Dv^2 \equiv 0 \pmod{p}$, откуда $D \equiv (uv^{-1})^2 \pmod{p}$. Для доказательства первого утверждения остаётся заметить, что $D = df^2$.

Второе утверждение очевидным образом следует из $p \nmid D$.

Третье утверждение следует из первого в силу хорошо известного факта из теории квадратичных полей (например, [9, Предложения 13.1.3 и 13.1.4]).

Для доказательства четвёртого утверждения рассмотрим по модулю 2 равенство $4q = u^2 + |D|v^2$. Если D чётно, то u чётно, $\mathcal{O}_D = \mathbb{Z}[(D + \sqrt{D})/2] = \mathbb{Z}[(\sqrt{D})/2]$ и $(u + v\sqrt{D})/2 = u/2 + v\sqrt{D}/2 \in \mathcal{O}_D$. Если же D нечётно, то $u \equiv v \pmod{2}$, $\mathcal{O}_D = \mathbb{Z}[(D + \sqrt{D})/2] = \mathbb{Z}[(1 + \sqrt{D})/2]$ и $(u + v\sqrt{D})/2 = (u - v)/2 + v(1 + \sqrt{D})/2 \in \mathcal{O}_D$.

Заметим, что $\frac{u + v\sqrt{D}}{2}\mathcal{O} \cdot \frac{u - v\sqrt{D}}{2}\mathcal{O} = q\mathcal{O} = p^n\mathcal{O} = \mathfrak{p}^n\bar{\mathfrak{p}}$. Поскольку $p \nmid u$, то $\mathfrak{p}\bar{\mathfrak{p}} = p\mathcal{O} \nmid \frac{u \pm v\sqrt{D}}{2}\mathcal{O}$. Теперь последнее утверждение следует из единственности разложения на простые идеалы в кольце \mathcal{O} . ■

Перейдём к вопросам реализации.

Действия на первом этапе зависят от требований к q и порядку кривой.

Если q фиксировано, то будем осуществлять перебор целых D , удовлетворяющих (1). Для каждого D сначала проверим необходимое условие (6); если оно нарушено, перейдём к следующему D . Пусть D удовлетворяет (6). Применим алгоритм Корначиа [10], который решает уравнение $x^2 + |D|y^2 = m$, при $m = 4q$. Если решения не существует, перейдём к следующему D . Если решение найдено, то проверим, удовлетворяет ли $q + 1 \pm x$ требованиям на порядок кривой.

Если же q не фиксировано, то вместо метода, описанного в предыдущем абзаце, эффективнее фиксировать D , удовлетворяющее (1), после чего случайным образом выбирать \hat{u}, \hat{v} , вычислять q из (4) и $q + 1 \pm \hat{u}$ и проверять, что они удовлетворяют нужным условиям. В [11] и [12] предлагаются улучшения этого метода, по существу заключающиеся в том, что если указанные параметры выбирать не совсем случайно, то можно гарантировать отсутствие (или снизить вероятность существования) малых простых делителей у $q = (\hat{u}^2 + |D|\hat{v}^2)/4$ и $q + 1 \pm \hat{u}$. Так, если ограничиться нечётным p и требованием нечётности и простоты одного из чисел $q + 1 \pm u$ (как это делает [12]), то легко видеть, что $D \equiv 5 \pmod{8}$, \hat{u} и \hat{v} должны быть нечётными; в [12] предлагается начать с $\hat{u} = 210\hat{u}_0 + 1$, $\hat{v} = 210\hat{v}_0 + 105$, \hat{u}_0, \hat{v}_0 — случайным образом выбранные натуральные числа, после чего (если выбранные \hat{u} и \hat{v} не подошли) прибавлять к \hat{u} попеременно 106 и $104 = 210 - 106$. Заметим, что $210 = 105 \cdot 2 = 2 \cdot 3 \cdot 5 \cdot 7$. При таком выборе $(\hat{u}^2 + |D|\hat{v}^2)/4$ и одно из $q + 1 \pm u$ не делится на 2, 3, 5, 7. Метод из [11] использует больше малых простых делителей и более громоздкий, поэтому здесь не приводится. Сравнение производительности методов есть в [12].

На втором этапе вычисляем многочлен $H_D[j](x)$. Для этого перечислим все $h = |\mathcal{H}_D|$ приведённых форм, вычислим их корни τ_1, \dots, τ_h , найдём с достаточно высокой точностью значения $j(\tau_1), \dots, j(\tau_h)$ как комплексные числа. По ним приближённо вычислим коэффициенты многочлена $H_D[j](x)$. Если точность вычислений такова, что погрешность в коэффициентах меньше 0,5, то коэффициенты, являющиеся целыми числами, могут быть однозначно восстановлены по вычисленным приближениям.

Если M — числовое поле, $\mathfrak{C} \subset \mathcal{O}_M$ — простой идеал, $z \in \mathcal{O}_M$, то через $R_{\mathfrak{C}}(z)$ будем обозначать редукцию z по модулю идеала \mathfrak{C} (таким образом, $R_{\mathfrak{C}}$ — отображение из \mathcal{O}_M в конечное поле). Отображение $R_{\mathfrak{C}}$ также действует на многочленах из $\mathcal{O}_M[x]$ покоэффициентно.

Для реализации третьего этапа надо показать, что многочлен $R_{p\mathbb{Z}}(H_D[j](x))$ разлагается на линейные множители над \mathbb{F}_q , а также построить эллиптическую кривую по её j -инварианту.

По лемме 1 $p\mathcal{O} = \mathfrak{p}\bar{\mathfrak{p}}$. Пусть $\mathfrak{B} \subset \mathcal{O}_L$ — простой идеал, лежащий над \mathfrak{p} . Поскольку $\mathfrak{B} \cap \mathbb{Z} = \mathfrak{p} \cap \mathbb{Z} = p\mathbb{Z}$, то

$$R_{p\mathbb{Z}}(H_D[j](x)) = R_{\mathfrak{B}}(H_D[j](x)) = \prod_{i=1}^h (x - R_{\mathfrak{B}}(j(\mathfrak{a}_i))) \quad (7)$$

и многочлен $R_{p\mathbb{Z}}(H_D[j](x))$ раскладывается на линейные множители над полем $\mathcal{O}_L/\mathfrak{B}$. Таким образом, остаётся доказать следующую теорему.

Теорема 1.

$$\mathcal{O}_L/\mathfrak{B} \subset \mathbb{F}_q. \quad (8)$$

Доказательство. Через $\left(\frac{L/K}{\mathfrak{c}}\right)$, где \mathfrak{c} — простой идеал \mathcal{O}_L , лежащий над простым идеалом $\mathfrak{c} \subset \mathcal{O}$, неразветвлённым в L , будем обозначать символ Артина [5, §5] (который определён для любого расширения Галуа $K \subset L$, но здесь понадобится только для конкретных полей K и L , определённых выше), а именно такой единственный ([5, Lemma 5.19]) элемент $\sigma \in \text{Gal}(L/K)$, что $\sigma(\alpha) \equiv \alpha^{\text{Norm}(\mathfrak{c})} \pmod{\mathfrak{c}}$ для любого $\alpha \in \mathcal{O}_L$. Поскольку $\text{Gal}(L/K) \cong \mathcal{H}_D$ абелева, то символ Артина зависит только от \mathfrak{c} [5, Corollary 5.21] и для него правомерно обозначение $\left(\frac{L/K}{\mathfrak{c}}\right)$. Если $\mathfrak{b} = \mathfrak{c}_1^{s_1} \dots \mathfrak{c}_k^{s_k}$ — дробный идеал \mathcal{O} и все \mathfrak{c}_i — простые, неразветвлённые в L , то определим $\left(\frac{L/K}{\mathfrak{b}}\right) = \left(\frac{L/K}{\mathfrak{c}_1}\right)^{s_1} \dots \left(\frac{L/K}{\mathfrak{c}_k}\right)^{s_k}$. Отображение $\left(\frac{L/K}{\cdot}\right)$ задаёт гомоморфизм из группы таких дробных идеалов \mathcal{O} , в разложение которых на простые не входят простые идеалы, разветвлённые в L , в группу $\text{Gal}(L/K)$. Этот гомоморфизм называется отображением Артина.

Через $P_{K,\mathbb{Z}}(f)$ [5, §9] будем обозначать подгруппу дробных идеалов \mathcal{O} , порождённую главными идеалами вида $\alpha\mathcal{O}$, $\alpha \in \mathcal{O}$, $\alpha \equiv a \pmod{f\mathcal{O}}$ для некоторого $a \in \mathbb{Z}$, $\text{gcd}(a, f) = 1$. Согласно [5, §9], поле L классов кольца \mathcal{O}_D — это единственное абелево расширение K , такое, что

- все простые идеалы \mathcal{O} , разветвлённые в L , делят $f\mathcal{O}$ (и, следовательно, все идеалы из $P_{K,\mathbb{Z}}(f)$ неразветвлены в L : если $\alpha \equiv a \pmod{f\mathcal{O}}$, то $\text{gcd}(\alpha\mathcal{O}, f\mathcal{O}) = \text{gcd}(a\mathcal{O}, f\mathcal{O}) = \text{gcd}(a, f)\mathcal{O} = \mathcal{O}$, так что идеал $\alpha\mathcal{O}$ взаимно прост с $f\mathcal{O}$);
- ядро отображения Артина есть группа $P_{K,\mathbb{Z}}(f)$.

Положим $\hat{\pi} = (\hat{u} + \hat{v}\sqrt{D})/2$; из леммы 1 следует, что либо $\mathfrak{p}^n = \hat{\pi}\mathcal{O}$, либо $\mathfrak{p}^n = \bar{\hat{\pi}}\mathcal{O}$. В обоих случаях идеал \mathfrak{p}^n главный и лежит в $P_{K,\mathbb{Z}}(f)$ (поскольку $\hat{\pi} = (\hat{u} + \hat{v}\sqrt{D})/2 \equiv (\hat{u} - f\hat{v}d)/2 \pmod{f\mathcal{O}}$, $\bar{\hat{\pi}} = (\hat{u} - \hat{v}\sqrt{D})/2 \equiv (\hat{u} + f\hat{v}d)/2 \pmod{f\mathcal{O}}$, по определению $P_{K,\mathbb{Z}}(f)$, лемме 1 и равенству (4)), следовательно, лежит в ядре отображения Артина.

Отсюда получаем, что $\left(\frac{L/K}{\mathfrak{p}}\right)^n = Id$. Иными словами, n -кратное возведение в степень $\text{Norm}(\mathfrak{p}) = p$, то есть возведение в степень $p^n = q$, действует на $\mathcal{O}_L/\mathfrak{B}$ тривиально, что возможно только в том случае, когда $\mathcal{O}_L/\mathfrak{B} \subset \mathbb{F}_q$. ■

Используя формулы из [7, Proposition A.1.1], легко убедиться, что при $j \in \mathbb{F}_q$ или $j \in \mathbb{C}$ следующие кривые, определённые соответственно над \mathbb{F}_q или над \mathbb{C} , имеют j -инвариант, равный j :

- над полем характеристики, равной 0 или не меньшей 5, при $j \neq 0$ и $j \neq 1728$: $y^2 = x^3 + 3cx + 2c$, где $c = \frac{j}{1728 - j}$;
- над полем характеристики, равной 0 или не меньшей 5, при $j = 0$: $y^2 = x^3 + 1$;
- над полем характеристики, равной 0 или не меньшей 5, при $j = 1728$: $y^2 = x^3 + x$;
- над полем \mathbb{F}_q характеристики 2 при $j \in \mathbb{F}_q^*$: $y^2 + xy = x^3 + j^{-1}$;
- над полем \mathbb{F}_q характеристики 3 при $j \in \mathbb{F}_q^*$: $y^2 = x^3 + x^2 - j^{-1}$.

Неуказанные случаи $j = 0$ для полей характеристик 2 и 3 задают суперсингулярные кривые [7, Exercise 5.7, Theorem 4.1], поэтому нам не встретятся.

На четвёртом этапе утверждается, что построенная к этому моменту кривая E'' обладает комплексным умножением на \mathcal{O}_D (и, в частности, несуперсингулярна). Докажем это.

Из (7) и построения кривой E'' следует, что её j -инвариант равен $R_{\mathfrak{B}}(j(\mathfrak{a}))$, где \mathfrak{a} — некоторый собственный дробный идеал \mathcal{O}_D . Поскольку $j(\mathfrak{a}) \in \mathcal{O}_L$, то в силу [13, §4.3] существует некоторое конечное расширение L' поля L , кривая E' , определённая над L' , с j -инвариантом, равным $j(\mathfrak{a})$, и идеал $\mathfrak{B}' \subset \mathcal{O}_{L'}$, лежащий над \mathfrak{B} , редукция по которому уравнения кривой задаёт кривую без особых точек (уже над конечным полем).

Поскольку j -инвариант кривой E' равен сингулярному значению, то E' обладает комплексным умножением на \mathcal{O}_D . Поскольку j -инвариант редукции равен редукции j -инварианта (ибо j -инвариант есть рациональная функция от коэффициентов уравнения, задающего кривую), а $\mathfrak{B}' \cap \mathcal{O}_L = \mathfrak{B}$, то $R_{\mathfrak{B}'}(j(\mathfrak{a})) = R_{\mathfrak{B}}(j(\mathfrak{a}))$, т. е. j -инвариант редукции кривой E' по модулю \mathfrak{B}' равен j -инварианту кривой E'' . Поскольку две кривые изоморфны, если и только если их j -инварианты совпадают [7, Proposition III.1.4], то E'' изоморфна редукции E' . Теперь из леммы 1 и свойств редукции [8, Теорема 13.12] следует, что E'' несуперсингулярна и $\text{End}(E'') \cong \text{End}(E') \cong \mathcal{O}_D$.

Таким образом, на четвёртом этапе мы начинаем с кривой E'' , которая определена над \mathbb{F}_q и имеет комплексное умножение на \mathcal{O}_D . Как было показано, порядок кривой E'' равен $q + 1 - u$, где $4q = u^2 + |D|v^2$ и $\gcd(q, u) = 1$, но u, v необязательно совпадают с \hat{u}, \hat{v} . Пусть $\pi = (u + v\sqrt{D})/2 \in \mathcal{O}_D \subset \mathcal{O}$. В силу леммы 1 либо $\pi\mathcal{O} = \mathfrak{p}^n$, либо $\pi\mathcal{O} = \bar{\mathfrak{p}}^n$. То же самое верно и для $\hat{\pi}\mathcal{O}$. Следовательно, $\pi\mathcal{O} = \hat{\pi}\mathcal{O}$ либо $\pi\mathcal{O} = \bar{\hat{\pi}}\mathcal{O}$. Поскольку норма идеала $\pi\mathcal{O}_D \subset \mathcal{O}_D$ равна q и взаимно проста с кондуктором f порядка \mathcal{O}_D , то $\pi\mathcal{O}_D = \pi\mathcal{O} \cap \mathcal{O}_D$ [5, Proposition 7.20]. Аналогично $\hat{\pi}\mathcal{O}_D = \hat{\pi}\mathcal{O} \cap \mathcal{O}_D$. Таким образом, либо $\pi\mathcal{O}_D = \hat{\pi}\mathcal{O}_D$, либо $\pi\mathcal{O}_D = \bar{\hat{\pi}}\mathcal{O}_D$, т. е. в кольце \mathcal{O}_D число π ассоциировано либо с $\hat{\pi}$, либо с $\bar{\hat{\pi}}$.

Хорошо известно (см., например, [9, Предложение 13.1.5]), что единицы в кольце \mathcal{O}_D суть $\{\pm 1\}$ при $D \notin \{-3, -4\}$, $\{\pm 1, \pm\zeta_3, \pm\zeta_3^2\}$ при $D = -3$, где $\zeta_3 = e^{2\pi i/3} = (-1 + \sqrt{-3})/2$, и $\{\pm 1, \pm i\}$ при $D = -4$.

Если $D \notin \{-3, -4\}$, то $\pi = \pm\hat{\pi}$ либо $\pi = \pm\bar{\hat{\pi}}$, что соответствует $u = \pm\hat{u}$. Таким образом, в этом случае $|E''(\mathbb{F}_q)| = q + 1 \pm \hat{u}$. Если $|E''(\mathbb{F}_q)| = q + 1 - \hat{u}$, то кривая E'' является искомой, иначе следует сконструировать другую кривую по уравнению E'' следующим образом. Если $p \neq 2$, то нормальная форма Вейерштрасса уравнения кривой имеет вид $y^2 = f(x)$, где f — многочлен степени 3 со старшим коэффициентом 1, и тогда кривая $y^2 = c^3 f(x/c)$, где c — квадратичный невычет из \mathbb{F}_q , имеет нужный порядок (как легко видеть из формулы в [14]). Если же $p = 2$, то нормальная форма Вейерштрасса имеет вид $y^2 + xy = x^3 + a_2 x^2 + a_6$, и тогда кривая $y^2 + xy = x^3 + (a_2 + \gamma)x^2 + a_6$, где $\text{Tr}_{\mathbb{F}_q/\mathbb{F}_2} \gamma = 1$, имеет нужный порядок [14].

Если $D = -3$, то в соответствии с описанной выше процедурой можно вычислить $H_{-3}[j](x) = x$, единственный корень $j = 0$. Формула (6) в этом случае означает $\left(\frac{-3}{p}\right) = 1$, откуда $p \equiv 1 \pmod{3}$, в частности, $p > 3$. Любая кривая вида $y^2 = x^3 + b$, $b \neq 0$, имеет j -инвариант, равный нулю [7, Proposition A.1.1], и все кривые такого вида $\bar{\mathbb{F}}_q$ -изоморфны между собой (поскольку у них одинаковый j -инвариант).

Обозначим через χ единственный мультипликативный характер \mathbb{F}_q порядка 2 и положим $S_2(b) = \sum_{x \in \mathbb{F}_q} \chi(x^3 + b)$. Легко видеть, что порядок кривой $y^2 = x^3 + b$ равен $q + 1 + S_2(b)$. Поскольку $p \equiv 1 \pmod{3}$, то и $q \equiv 1 \pmod{3}$; в этом случае, соглас-

но [15] (где рассматривается только случай $q = p$, когда χ есть символ Лежандра, но рассуждения переносятся на общий случай тривиальным образом), существуют $k, l \in \mathbb{Z}$, такие, что для произвольного кубического невычета $c \in \mathbb{F}_q^*$ справедливы равенства $S_2(1) = 2k$, $S_2(c^2) = -k \pm 3l$, $S_2(c^{-2}) = -k \mp 3l$ и $q = k^2 + 3l^2$. Кроме того, $S_2(b) = \chi(t)S_2(bt^3)$ для любого $t \in \mathbb{F}_q^*$.

Построенная на третьем этапе кривая E'' имеет вид $y^2 = x^3 + 1$; тогда $u = -S_2(1) = -2k$, $q = k^2 + 3l^2$, $l = \pm \frac{S_2(c^2) - S_2(c^{-2})}{6}$, где c — произвольный кубический невычет в \mathbb{F}_q ; отсюда в силу $|\pi|^2 = q$ имеем $\pi = -k \pm l\sqrt{-3}$. В случае необходимости сделаем замену c на c^{-1} и представим π в виде $-k - l\sqrt{-3}$, где $l = (S_2(c^2) - S_2(c^{-2}))/6$.

Либо $\hat{\pi}$, либо $\bar{\hat{\pi}}$ равно произведению π на какую-то из единиц кольца \mathcal{O}_{-3} ; в обоих случаях $\hat{u} = 2 \operatorname{Re} \hat{\pi}$ равно удвоенной вещественной части произведения π на какую-то из единиц. Таким образом, для \hat{u} возможны шесть вариантов:

- $\hat{u} = 2 \operatorname{Re} \hat{\pi} = \pm 2 \operatorname{Re} \pi = \pm 2k$. В этом случае ищем кривую с $q + 1 \pm 2k$ точек, и кривая $y^2 = x^3 + 1$ либо $y^2 = x^3 + g^3$, где g — квадратичный невычет в \mathbb{F}_q , является искомой;
- $\hat{u} = 2 \operatorname{Re} \hat{\pi} = \pm 2 \operatorname{Re} (\zeta_3 \pi) = \pm (k + 3l)$. В этом случае ищем кривую с $q + 1 \pm (k + 3l)$ точек, и кривая $y^2 = x^3 + c^2$ либо $y^2 = x^3 + c^2 g^3$ является искомой;
- $\hat{u} = 2 \operatorname{Re} \hat{\pi} = \pm 2 \operatorname{Re} (\zeta_3^2 \pi) = \pm (k - 3l)$. В этом случае ищем кривую с $q + 1 \pm (k - 3l)$ точек, и кривая $y^2 = x^3 + c^{-2}$ либо $y^2 = x^3 + c^{-2} g^3$ является искомой.

Если $D = -4$, то аналогично имеем $H_{-4}[j](x) = x - 1728$, единственный корень $j = 1728$. Формула (6) в этом случае означает $\left(\frac{-4}{p}\right) = 1$, откуда $p \equiv 1 \pmod{4}$, в частности по-прежнему $p > 3$. Любая кривая вида $y^2 = x^3 + bx$, $b \neq 0$, имеет j -инвариант, равный 1728 [7, Proposition A.1.1], и все кривые такого вида \mathbb{F}_q -изоморфны между собой (поскольку имеют одинаковый j -инвариант).

Положим $S_1(b) = \sum_{x \in \mathbb{F}_q} \chi(x)\chi(x^2 + b)$, где, как и раньше, χ — единственный мультипликативный характер \mathbb{F}_q порядка 2. Легко видеть, что порядок кривой $y^2 = x^3 + bx$ равен $q + 1 + S_1(b)$. Поскольку $p \equiv 1 \pmod{4}$, то и $q \equiv 1 \pmod{4}$; в этом случае, согласно [16], существуют $k, l \in \mathbb{Z}$, такие, что k нечётно, $S_1(1) = 2k$, $S_1(b) = \pm 2l$ для любого квадратичного невычета b и $S_1(b) = \chi(t)S_1(bt^2)$ для любого $t \in \mathbb{F}_q^*$.

Построенная на третьем этапе кривая E'' имеет вид $y^2 = x^3 + x$; тогда $u = -S_1(1) = -2k$, $q = k^2 + l^2$. Поскольку $|\pi|^2 = q$, то $\pi = -k \pm li$.

Аналогично предыдущему случаю есть четыре варианта для \hat{u} , а именно $\pm 2 \operatorname{Re} \pi = \pm 2k$ и $\pm 2 \operatorname{Re} (i\pi) = \pm 2l$. Если $\hat{u} = \pm 2k$, то кривая $y^2 = x^3 + x$ либо $y^2 = x^3 + g^2 x$, где g — квадратичный невычет в \mathbb{F}_q , имеет нужное число точек. Если $\hat{u} = \pm 2l$, то кривая $y^2 = x^3 + gx$ либо $y^2 = x^3 + g^3 x$ имеет нужное число точек.

1.3. Некоторые известные оптимизации метода

Коэффициенты многочлена $H_D[j]$ растут достаточно быстро с ростом $|D|$. Например, $H_{-40}[j](x) = x^2 - 425692800x + 9103145472000$. Поэтому представляет интерес поиск других функций, сингулярные значения которых лежат в L , но имеют меньшую высоту характеристического многочлена.

Пусть $z \in \mathbb{H}$, $q = e^{2\pi iz}$. Введём вслед за [6] функции

$$\begin{aligned} \eta(z) &= q^{\frac{1}{24}} \prod_{n=1}^{\infty} (1 - q^n) = q^{\frac{1}{24}} \sum_{n=-\infty}^{\infty} (-1)^n q^{\frac{3n^2+n}{2}}, \\ \mathfrak{f}(z) &= e^{-\frac{\pi i}{24}} \frac{\eta\left(\frac{z+1}{2}\right)}{\eta(z)}, \quad \mathfrak{f}_1(z) = \frac{\eta\left(\frac{z}{2}\right)}{\eta(z)}, \quad \mathfrak{f}_2(z) = \sqrt{2} \frac{\eta(2z)}{\eta(z)}, \\ \gamma_2(z) &= \frac{\mathfrak{f}^{24} - 16}{\mathfrak{f}^8} = \frac{\mathfrak{f}_1^{24} + 16}{\mathfrak{f}_1^8} = \frac{\mathfrak{f}_2^{24} + 16}{\mathfrak{f}_2^8}, \quad j(z) = \gamma_2(z)^3. \end{aligned} \quad (9)$$

Пусть N — натуральное число. Назовём вслед за [17] N -системой набор форм $(A_1, B_1, C_1), \dots, (A_h, B_h, C_h)$, такой, что

- набор $\{\mathfrak{h}(A_i, B_i, C_i) : 1 \leq i \leq h\}$ является полной системой представителей группы \mathcal{H}_D ;
- справедливы соотношения $\gcd(A_i, N) = 1$; $B_i \equiv B_j \pmod{2N}$.

Отметим, что для любой формы (A_i, B_i, C_i) выполнено $B_i \equiv D \pmod{2}$, так что из первого условия автоматически следует, что $B_i \equiv B_j \pmod{2}$ для любых i, j .

По известному набору форм, удовлетворяющему первому условию (например, полному набору приведённых форм), можно построить N -систему. Соответствующий алгоритм приведён в [17, доказательство Proposition 3]. (Предполагаем известным разложение N на простые множители.)

1) Сначала добьёмся выполнения условия $\gcd(A_i, N) = 1$ при всех i .

Ясно, что достаточно уметь добиваться условия $\gcd(A_i, N_0 l) = 1$ в предположении, что $\gcd(A_i, N_0) = 1$, где l — очередной простой делитель N , не делящий N_0 .

Число l не может делить каждое из трёх чисел A_i , $A_i + N_0 B_i + N_0^2 C_i$, $l^2 A_i + l N_0 B_i + N_0^2 C_i$, потому что иначе числа A_i, B_i, C_i имели бы общий делитель l и форма (A_i, B_i, C_i) не была бы примитивной.

- Если $l \nmid A_i$, то условие $\gcd(A_i, N_0 l) = 1$ выполнено.
- Если $l \nmid A_i + N_0 B_i + N_0^2 C_i$, то в форме $A_i x^2 + B_i x y + C_i y^2$ сделаем замену переменных $x = x'$, $y = N_0 x' + y'$ (очевидно, определитель матрицы этой замены равен 1); получим эквивалентную форму с коэффициентом при x'^2 , равным $A_i + N_0 B_i + N_0^2 C_i$, дальше будем работать с ней вместо (A_i, B_i, C_i) .
- Если $l \nmid l^2 A_i + l N_0 B_i + N_0^2 C_i$, то найдём $a, b \in \mathbb{Z}$, такие, что $al - bN_0 = 1$, и сделаем замену переменных $x = lx' + by'$, $y = N_0 x' + ay'$ (очевидно, определитель матрицы этой замены равен 1); получим эквивалентную форму с коэффициентом при x'^2 , равным $l^2 A_i + l N_0 B_i + N_0^2 C_i$, дальше будем работать с ней вместо (A_i, B_i, C_i) .

2) Остаётся обеспечить условие $B_i \equiv B_1 \pmod{2N}$ при всех i . Замена переменных $x = x' + ay'$, $y = y'$ переводит форму (A_i, B_i, C_i) в эквивалентную форму $(A_i, B_i + 2aA_i, C_i + aB_i + a^2 A_i)$; поскольку $\gcd(A_i, N) = 1$, то применение этого преобразования с $a = A_i^{-1}(B_1 - B_i)/2 \pmod{N}$ решает задачу.

Теорема 2 (Theorem 1 из [17]). Пусть $\alpha \in \mathbb{H}$ — корень формы

$$(A, B, C), \quad 2 \nmid A, \quad 32 \mid B,$$

дискриминант которой равен $B^2 - 4AC = D = -4m$, $m \in \mathbb{N}$. Пусть $g(\alpha)$ определено следующими формулами:

$$\left(\left(\frac{2}{A} \right) \frac{1}{\sqrt{2}} \mathfrak{f}(\alpha^2) \right)^3, \quad \text{если } m \equiv 1 \pmod{8},$$

$$\begin{aligned}
 & f(\alpha)^3, \text{ если } m \equiv 3 \pmod{8}, \\
 & \left(\frac{1}{2}f(\alpha)^4\right)^3, \text{ если } m \equiv 5 \pmod{8}, \\
 & \left(\left(\frac{2}{A}\right)\frac{1}{\sqrt{2}}f(\alpha)\right)^3, \text{ если } m \equiv 7 \pmod{8}, \\
 & \left(\left(\frac{2}{A}\right)\frac{1}{\sqrt{2}}f_1(\alpha)^2\right)^3, \text{ если } m \equiv 2 \pmod{4}, \\
 & \left(\left(\frac{2}{A}\right)\frac{1}{2\sqrt{2}}f_1(\alpha)^4\right)^3, \text{ если } m \equiv 4 \pmod{8}.
 \end{aligned}$$

Тогда $g(\alpha) \in \mathcal{O}_L$.

Если $\alpha_1 = \alpha, \dots, \alpha_h$ — корни элементов 16-системы, то сингулярные значения $g(\alpha_i)$ образуют полный набор различных сопряжённых чисел над \mathbb{Q} .

Теорема 3 (Theorem 2 из [17]). Пусть $\alpha \in \mathbb{H}$ — корень формы

$$(A, B, C), \quad 3 \nmid A, \quad 3 \mid B$$

дискриминанта $B^2 - 4AC = D$. Тогда

$$\mathbb{Q}(\gamma_2(\alpha)) = \begin{cases} \mathbb{Q}(j(\alpha)), & 3 \nmid D, \\ \mathbb{Q}(j(3\alpha)), & 3 \mid D. \end{cases}$$

Более того, в случае $3 \nmid D$, если $\alpha_1 = \alpha, \dots, \alpha_h$ — корни элементов 3-системы, то сингулярные значения $\gamma_2(\alpha_i)$ образуют полный набор различных сопряжённых чисел над \mathbb{Q} . Кроме того, числа $\gamma_2(\alpha_i)$ целые алгебраические.

Введём вслед за [18] при простых p_1, p_2 функцию

$$\mathbf{m}_{p_1, p_2}(z) = \frac{\eta\left(\frac{z}{p_1}\right)\eta\left(\frac{z}{p_2}\right)}{\eta(z)\eta\left(\frac{z}{p_1 p_2}\right)}$$

и обозначим $s = \frac{24}{\gcd(24, (p_1 - 1)(p_2 - 1))}$.

Теорема 4 (Theorems 3.2, 3.3, Corollary 3.1 из [18]). Пусть D удовлетворяет (1), $N = p_1 p_2$, p_1 и p_2 — простые числа, удовлетворяющие следующему условию:

- 1) $\left(\frac{D}{p_1}\right), \left(\frac{D}{p_2}\right) \neq -1$ при $p_1 \neq p_2$,
- либо 2a) $\left(\frac{D}{p}\right) = 1$ при $p_1 = p_2 = p$, либо 2б) $p \mid f$ при $p_1 = p_2 = p$.

Тогда существует форма (A_1, B_1, C_1) , такая, что $\gcd(A_1, N) = 1$ и $N \mid C_1$. Пусть $\alpha_1 \in \mathbb{H}$ — её корень. Сингулярное значение $\mathbf{m}_{p_1, p_2}^s(\alpha_1)$ лежит в L . Все сопряжённые над K значения к $\mathbf{m}_{p_1, p_2}^s(\alpha_1)$ суть $\mathbf{m}_{p_1, p_2}^s(\alpha_i)$, где α_i пробегает корни элементов N -системы. Числа $\mathbf{m}_{p_1, p_2}^s(\alpha_i)$ являются целыми алгебраическими.

Если выполнено условие 1 или 2a), то $\mathbf{m}_{p_1, p_2}^s(\alpha_i)$ являются единицами (т. е. $\mathbf{m}_{p_1, p_2}^{-s}(\alpha_i)$ также целые алгебраические).

Если p_1 и p_2 удовлетворяют более сильному условию

$$- \left(\frac{D}{p_1}\right), \left(\frac{D}{p_2}\right) \neq -1 \text{ и } p_1, p_2 \nmid f \text{ при } p_1 \neq p_2;$$

- $\left(\frac{D}{p}\right) = 1$ или $p \mid f$ при $p_1 = p_2 = p \neq 2$;
- либо $\left(\frac{D}{2}\right) = 1$, либо $2 \mid f$ и $D \not\equiv 4 \pmod{32}$ при $p_1 = p_2 = 2$,

то комплексное сопряжение переставляет числа $\mathbf{m}_{p_1, p_2}^s(\alpha_i)$.

Далее понадобятся более точные утверждения. Формулировки теорем 2, 3, 4 дают неполную информацию о действии автоморфизмов из $\text{Gal}(L/K)$ на сингулярных значениях. Однако их доказательства из [17, 18] предоставляют эту информацию.

Утверждение 1 [17]. Пусть θ — функция, описанная в формулировке одной из теорем 2, 3, 4 (в последней достаточно выполнения слабого условия на p_1, p_2), \mathbf{a}, \mathbf{b} — два элемента N -системы из формулировок тех же теорем, α — корень \mathbf{a} , β — корень \mathbf{b} , $\Omega : \mathcal{H}_D \rightarrow \text{Gal}(L/K)$ — канонический изоморфизм. Тогда

$$\theta(\alpha)^{\Omega(\mathfrak{h}(\mathbf{a})\mathfrak{h}(\mathbf{b})^{-1})} = \theta(\beta).$$

Для $\theta = j$ эта формула также верна, что уже упоминалось выше.

Утверждение 1 удобнее использовать в виде формулы, задающей действие конкретного элемента $\text{Gal}(L/K)$ на сингулярных значениях. Напомним, что \mathfrak{h} сюръективно, а N -система по определению содержит представителей всех классов \mathcal{H}_D .

Следствие. Пусть θ , N -система и \mathbf{a} такие же, как в предыдущем утверждении, и $\mathbf{c} \in \mathcal{H}_D$. Тогда существует форма \mathbf{b} , принадлежащая этой N -системе и такая, что

$$\mathfrak{h}(\mathbf{b}) = \mathfrak{h}(\mathbf{a})\mathbf{c}^{-1}.$$

Если β — корень \mathbf{b} , то

$$\theta(\alpha)^{\Omega(\mathbf{c})} = \theta(\beta). \quad (10)$$

В дальнейшем при использовании функции \mathbf{m}_{p_1, p_2}^s будем предполагать, что p_1 и p_2 удовлетворяют сильному условию теоремы; легко видеть, что для любого D такие p_1, p_2 всегда можно найти.

Объединяя теоремы 2 и 3, получаем, что если в дополнение к условиям теоремы 2 выполнены условия $3 \nmid D$, $3 \nmid A$, $3 \mid B$, то в заключении теоремы 2 в выражениях для $g(\alpha)$ можно убрать возведение в куб. Например, рассмотрим случай $m \equiv 3 \pmod{8}$. Согласно (9),

$$\mathfrak{f}(\alpha) = \frac{(\mathfrak{f}(\alpha)^3)^3 \gamma_2(\alpha)}{(\mathfrak{f}(\alpha)^3)^8 - 16}. \quad (11)$$

Поскольку $\mathfrak{f}(\alpha)^3 \in L$ и $\gamma_2(\alpha) \in L$, то и $\mathfrak{f}(\alpha) \in L$. Из утверждения 1 следует, что любой автоморфизм из $\text{Gal}(L/K)$ переводит $\mathfrak{f}(\alpha)^3$ в $\mathfrak{f}(\alpha')^3$ и $\gamma_2(\alpha)$ в $\gamma_2(\alpha')$, где α' зависит только от автоморфизма; из (11) следует, что $\mathfrak{f}(\alpha)$ переходит в $\mathfrak{f}(\alpha')$. Наконец, $\mathfrak{f}(\alpha)$ — целое алгебраическое, например, как кубический корень из целого алгебраического числа $\mathfrak{f}(\alpha)^3$. В остальных случаях формулы несколько сложнее, но принцип одинаков.

Пусть θ и $\alpha_* = \{\alpha_1, \dots, \alpha_h\}$ — соответственно функция и набор корней, заданные в условии одной из теорем 2–4. Введём многочлен от одной переменной

$$H_D[\theta, \alpha_*](x) = \prod_{i=1}^h (x - \theta(\alpha_i)).$$

Если θ — функция, заданная в условии теоремы 2 или 3, то многочлен $H_D[\theta, \alpha_*]$ имеет целые коэффициенты. Если же $\theta = \mathbf{m}_{p_1, p_2}^s$, то нужно привлечь утверждение 1, из

которого легко видеть, что $H_D[\theta, \alpha_*]$ инвариантен относительно $\text{Gal}(L/K)$ и, следовательно, лежит в $K[x]$; теперь из теоремы 4 следует, что и в этом случае $H_D[\theta, \alpha_*]$ имеет целые коэффициенты.

Например, $H_{-40}[\gamma_2, \alpha_*](x) = x^2 - 780x + 20880$, $H_{-40}[g, \alpha_*](x) = x^2 - x - 1$, где $g(\alpha) = \left(\frac{2}{A}\right) \frac{1}{\sqrt{2}} f_1(\alpha)^2$, $H_{-40}[\mathbf{m}_{5,7}, \alpha_*](x) = x^2 - x - 1$, $H_{-40}[\mathbf{m}_{11,13}, \alpha_*](x) = x^2 \pm 2x + 1$. (В первых трёх примерах многочлен не зависит от набора α_* , в последнем в зависимости от α_* есть два варианта.) Последний пример демонстрирует, что значения $\mathbf{m}_{p_1, p_2}^s(\alpha_i)$ могут и не быть попарно различными, так что в общем случае $H_D[\mathbf{m}_{p_1, p_2}^s, \alpha_*]$ есть некоторая степень минимального многочлена.

Поскольку для рассматриваемых функций многочлен $H_D[\theta, \alpha_*]$, как и $H_D[j]$, имеет целые коэффициенты, то его можно вычислять путём построения достаточно точных приближений к сингулярным значениям $\theta(\alpha_i)$ с последующим перемножением скобок вида $x - \theta(\alpha_i)$ и восстановлением целых коэффициентов многочлена путём округления вычисленных приближений. Поскольку $\theta(\alpha_i) \in \mathcal{O}_L$, то $\theta(\alpha_i)$ имеет представителя в $\mathcal{O}_L/\mathfrak{B} \subset \mathbb{F}_q$ (теорема 1), так что редукция уравнения $H_D[\theta, \alpha_*](x) = 0$ по модулю p разлагается на линейные множители над \mathbb{F}_q . Остаётся по значению редукции $\theta(\alpha_i)$ в \mathbb{F}_q вычислить j -инвариант нужной эллиптической кривой, равный $R_{\mathfrak{B}}(j(\alpha))$. Для функции γ_2 и степеней f из теоремы 2 ответ дают формулы (9). Для функций \mathbf{m}_{p_1, p_2}^s ситуация более сложная. В этом случае существует многочлен $\Phi_{p_1, p_2}(x, y) \in \mathbb{Z}[x, y]$, такой, что выполнено тождество $\Phi_{p_1, p_2}(\mathbf{m}_{p_1, p_2}^s(z), j(z)) = 0$ [18]; подставляя в это тождество значение $z = \alpha_i$ и редуцируя по модулю \mathfrak{B} (поскольку $\mathfrak{B} \cap \mathbb{Z} = p\mathbb{Z}$, то для редукции достаточно привести многочлен Φ_{p_1, p_2} по модулю p), получаем полиномиальное уравнение над \mathbb{F}_q на $R_{\mathfrak{B}}(j(\alpha))$, из которого можно найти несколько вариантов для $R_{\mathfrak{B}}(j(\alpha))$. Правильный вариант можно выбрать, рассмотрев все варианты, построив для каждой эллиптической кривую и проверив, что её порядок равен $q + 1 - m$. Если, как, например, нужно в криптографических приложениях, $q + 1 - m$ делится на большое простое число, то простой тест, проверяющий, что случайно выбранная точка кривой после умножения на $q + 1 - m$ переходит в нуль, хорошо отсеивает неверные альтернативы. Следует отметить, что совпадающий порядок кривой не гарантирует, что кольцо эндоморфизмов есть в точности \mathcal{O}_D , но столь тонкие различия часто неважны для приложений; более подробно этот вопрос разобран в [18].

2. Свойства изоморфизма Ω

Мы определили группу \mathcal{H}_D как фактор-группу группы $I(\mathcal{O}_D)$ собственных дробных идеалов порядка \mathcal{O}_D по подгруппе главных идеалов $P(\mathcal{O}_D)$.

Напомним, что идеал $\mathfrak{a} \subset \mathcal{O}_D$ взаимно прост с f , если $\mathfrak{a} + f\mathcal{O}_D = \mathcal{O}_D$. Согласно [5, Lemma 7.18], это условие эквивалентно $\text{gcd}(\text{Norm}(\mathfrak{a}), f) = 1$, и все такие идеалы собственные. Обозначим подгруппу, порождённую ими в $I(\mathcal{O}_D)$, через $I(\mathcal{O}_D, f)$. Обозначим подгруппу в $P(\mathcal{O}_D)$, порождённую главными идеалами $\alpha\mathcal{O}_D$ с $\text{gcd}(\text{Norm}(\alpha), f) = 1$, через $P(\mathcal{O}_D, f)$. Включение $I(\mathcal{O}_D, f) \subset I(\mathcal{O}_D)$ индуцирует изоморфизм $I(\mathcal{O}_D, f)/P(\mathcal{O}_D, f) \cong \mathcal{H}_D$ [5, Proposition 7.19].

Идеал $\mathfrak{a} \subset \mathcal{O}$ взаимно прост с f тогда и только тогда, когда выполнено условие $\text{gcd}(\text{Norm}(\mathfrak{a}), f) = 1$ [5, Lemma 7.18]. Обозначим группу дробных идеалов \mathcal{O} , порождённую такими идеалами, через $I(\mathcal{O}, f)$. Напомним, что $P_{K, \mathbb{Z}}(f)$ обозначает подгруппу идеалов \mathcal{O} , порождённую главными идеалами вида $\alpha\mathcal{O}$, где $\alpha \in \mathcal{O}$, $\alpha \equiv a \pmod{f\mathcal{O}}$ для $a \in \mathbb{Z}$, такого, что $\text{gcd}(a, f) = 1$. В соответствии с [5, Proposition 7.20] формула $\mathfrak{a} \mapsto \mathfrak{a}\mathcal{O}$

задаёт изоморфизм групп $\Omega_1 : I(\mathcal{O}_D, f) \rightarrow I(\mathcal{O}, f)$, сохраняющий норму. Кроме того [5, Proposition 7.22], Ω_1 индуцирует изоморфизм $I(\mathcal{O}_D, f)/P(\mathcal{O}_D, f) \cong I(\mathcal{O}, f)/P_{K,\mathbb{Z}}(f)$.

Таким образом, получили изоморфизм $\Omega_2 : \mathcal{H}_D \rightarrow I(\mathcal{O}, f)/P_{K,\mathbb{Z}}(f)$. Отображение Артина $I(\mathcal{O}, f) \rightarrow \text{Gal}(L/K)$, которое будем обозначать $\left(\frac{L/K}{\cdot}\right)$, индуцирует изоморфизм $I(\mathcal{O}, f)/P_{K,\mathbb{Z}}(f) \rightarrow \text{Gal}(L/K)$, композиция которого с Ω_2 , согласно [5, §9], есть канонический изоморфизм Ω , который фигурирует в утверждении 1.

Подытожим: существует коммутативная диаграмма

$$\begin{array}{ccccccc} I(\mathcal{O}_D) & \supset & I(\mathcal{O}_D, f) & \xrightarrow{\Omega_1} & I(\mathcal{O}, f) & \searrow & \left(\frac{L/K}{\cdot}\right) \\ \downarrow & & \downarrow & & \downarrow & & \text{Gal}(L/K), \\ \mathcal{H}_D & \longrightarrow & I(\mathcal{O}_D, f)/P(\mathcal{O}_D, f) & \longrightarrow & I(\mathcal{O}, f)/P_{K,\mathbb{Z}}(f) & \longrightarrow & \\ & \searrow & \xrightarrow{\Omega_2} & \xrightarrow{\Omega} & & & \end{array} \quad (12)$$

где вертикальные стрелки обозначают проекции группы на фактор-группу, а стрелки во второй строке обозначают изоморфизмы.

Теорема 5. Пусть (A, B, C) — форма, такая, что $\gcd(A, D) = 1$. Пусть $q \mid D$ такое, что выполнен один из двух наборов условий:

- $|q|$ — нечётное простое число, $q \equiv 1 \pmod{4}$;
- $q \in \{-4, \pm 8\}$, $D/q \equiv 0 \pmod{4}$ или $D/q \equiv 1 \pmod{4}$.

Тогда

- 1) $\sqrt{q} \in L$;
- 2) $\mathfrak{a} = \langle A, (-B + \sqrt{D})/2 \rangle_{\mathbb{Z}} \in I(\mathcal{O}_D, f)$, $\text{Norm}(\mathfrak{a}) = A$;
- 3) $\left(\frac{L/K}{\Omega_1(\mathfrak{a})}\right)(\sqrt{q}) = \left(\frac{q}{A}\right)\sqrt{q}$, где $\left(\frac{L/K}{\cdot}\right)$ обозначает отображение Артина, введённое при доказательстве теоремы 1, а $\left(\frac{q}{A}\right)$ есть символ Кронекера, определённый на с. 21.

Доказательство. Первое утверждение следует из [19, Theorem 2.2.23 и (2.2.8)].

По [5, Theorem 7.7] \mathfrak{a} является собственным идеалом \mathcal{O}_D . Его норма по определению равна $|\mathcal{O}_D/\mathfrak{a}|$; легко видеть, что в каждом смежном классе по \mathfrak{a} существует единственное целое число из $0, \dots, A-1$, так что $\text{Norm}(\mathfrak{a}) = A$. Поскольку $\gcd(A, f) = 1$, то \mathfrak{a} взаимно прост с f .

Докажем третье утверждение. Запишем $\Omega_1(\mathfrak{a}) = \mathfrak{p}_1 \cdot \dots \cdot \mathfrak{p}_s$, где \mathfrak{p}_i — простые идеалы \mathcal{O} (не обязательно различные). Поскольку $A = \text{Norm}(\mathfrak{a}) = \text{Norm}(\mathfrak{p}_1) \cdot \dots \cdot \text{Norm}(\mathfrak{p}_s)$ и символ Кронекера мультипликативен по нижнему аргументу, достаточно доказать, что для каждого простого $\mathfrak{p} \mid \Omega_1(\mathfrak{a})$ выполнено равенство с символом Артина

$$\left(\frac{L/K}{\mathfrak{p}}\right)(\sqrt{q}) = \left(\frac{q}{\text{Norm}(\mathfrak{p})}\right)\sqrt{q}. \quad (13)$$

Поскольку левая часть есть образ \sqrt{q} под действием некоторого автоморфизма, то она должна быть равна $\pm\sqrt{q}$.

Пусть \mathfrak{p} — простой идеал, $\mathfrak{p} \mid \Omega_1(\mathfrak{a})$, $\mathfrak{p} \cap \mathbb{Z} = p\mathbb{Z}$, p нечётно. Пусть простой идеал $\mathfrak{B} \subset \mathcal{O}_L$ лежит над \mathfrak{p} . Поскольку $\gcd(A, D) = 1$ и $q \mid D$, то $2\sqrt{q} \notin \mathfrak{B}$ и, следовательно, $\sqrt{q} \not\equiv -\sqrt{q} \pmod{\mathfrak{B}}$. По определению

$$\left(\frac{L/K}{\mathfrak{p}}\right)(\sqrt{q}) \equiv \sqrt{q}^{\text{Norm}(\mathfrak{p})} = q^{\frac{\text{Norm}(\mathfrak{p})-1}{2}}\sqrt{q} \pmod{\mathfrak{B}}.$$

Если идеал $p\mathcal{O}$ простой (т. е. $\mathfrak{p} = p\mathcal{O}$), то $\text{Norm}(\mathfrak{p}) = p^2$ и правая часть (13) равна \sqrt{q} . С другой стороны, $q^{\frac{\text{Norm}(\mathfrak{p})-1}{2}} = (q^{p-1})^{\frac{p+1}{2}} \equiv 1 \pmod{p}$, так что левая часть (13) сравнима с \sqrt{q} по модулю \mathfrak{B} и, следовательно, равна \sqrt{q} , так что равенство (13) выполнено.

Если идеал $p\mathcal{O}$ не простой, то $\text{Norm}(\mathfrak{p}) = p$ и правая часть (13) равна $\left(\frac{q}{p}\right)\sqrt{q}$. С другой стороны, $q^{\frac{\text{Norm}(\mathfrak{p})-1}{2}} = q^{\frac{p-1}{2}} \equiv \left(\frac{q}{p}\right) \pmod{p}$, так что левая часть (13) сравнима с $\left(\frac{q}{p}\right)\sqrt{q}$ по модулю \mathfrak{B} , следовательно, равна $\left(\frac{q}{p}\right)\sqrt{q}$, что доказывает (13) и в этом случае.

Пусть теперь $\mathfrak{p} \mid \Omega_1(\mathfrak{a})$, $\mathfrak{p} \cap \mathbb{Z} = 2\mathbb{Z}$, простой идеал $\mathfrak{B} \subset \mathcal{O}_L$ лежит над \mathfrak{p} . В этом случае $2 \mid A$, следовательно, по условию $2 \nmid D$ и q нечётно. Поскольку $B^2 - 4AC = D$, то $D \equiv B^2 \equiv 1 \pmod{8}$. Следовательно, $d \equiv 1 \pmod{8}$ и идеал $2\mathcal{O}$ не простой [9, Предложение 13.1.4], так что $\text{Norm}(\mathfrak{p}) = 2$. Таким образом, правая часть (13) равна $\left(\frac{q}{2}\right)\sqrt{q}$. Для вычисления левой части (13) рассмотрим $\left(\frac{L/K}{\mathfrak{p}}\right)\left(\frac{1+\sqrt{q}}{2}\right)$. Это выражение должно быть равно $(1 \pm \sqrt{q})/2$, и два возможных значения различны по модулю \mathfrak{B} . По определению

$$\left(\frac{L/K}{\mathfrak{p}}\right)\left(\frac{1+\sqrt{q}}{2}\right) \equiv \left(\frac{1+\sqrt{q}}{2}\right)^2 = \frac{q-1}{4} + \frac{1+\sqrt{q}}{2} \pmod{\mathfrak{B}}.$$

Если $\left(\frac{q}{2}\right) = 1$, то $q \equiv 1 \pmod{8}$, $(q-1)/4$ чётно и, следовательно, лежит в \mathfrak{B} ; если $\left(\frac{q}{2}\right) = -1$, то $q \equiv 5 \pmod{8}$, $(q-1)/4$ нечётно и, следовательно, сравнимо с $-1 \equiv 1$ по модулю \mathfrak{B} . В обоих случаях

$$\left(\frac{L/K}{\mathfrak{p}}\right)\left(\frac{1+\sqrt{q}}{2}\right) \equiv \frac{1+\left(\frac{q}{2}\right)\sqrt{q}}{2} \pmod{\mathfrak{B}},$$

откуда следует (13). ■

Лемма 2. Пусть $d < 0$ удовлетворяет одному из условий (2) и (3). Тогда d может быть единственным с точностью до порядка множителей способом представлен в виде произведения $d = q_1^* \cdot \dots \cdot q_t^*$, где все q_i^* попарно взаимно просты, $q^* = (-1)^{\frac{q-1}{2}}q$, если $q > 0$ — нечётное простое, и $q^* \in \{-4, \pm 8\}$, если $q = 2$.

Доказательство. Единственность такого представления очевидна, нужно доказать его существование.

В случае (2) разложение числа d на простые множители имеет вид $d = -q_1 \cdot \dots \cdot q_t$, где q_i — различные нечётные простые; поскольку $q_i^* = \pm q_i$, то $d = \pm q_1^* \cdot \dots \cdot q_t^*$; наконец, поскольку $d \equiv 1 \pmod{4}$ и $q_i^* \equiv 1 \pmod{4}$ для всех i , то заключение леммы выполнено.

В случае (3) разложение числа $d/4$ на простые множители имеет вид либо $d/4 = -q_1 \cdot \dots \cdot q_{t-1}$, либо $d/4 = -2q_1 \cdot \dots \cdot q_{t-1}$, где в обоих вариантах q_i — различные нечётные простые. Если $d/4$ нечётно, то, как и в предыдущем случае, $d/4 = \pm q_1^* \cdot \dots \cdot q_{t-1}^*$, но теперь в силу (3) $d/4 \not\equiv 1 \pmod{4}$, так что в правой части должен быть знак минус, и после умножения на 4 получаем заключение леммы с $q_t^* = -4$. Наконец, если $d/4$ чётно, то $d/4 = \pm 2q_1^* \cdot \dots \cdot q_{t-1}^*$, и, выбирая знак $q_t^* = \pm 8$ подходящим образом, получаем заключение леммы. ■

Легко видеть, что числа q_i^* из леммы 2 удовлетворяют условию теоремы 5 и, следовательно, $K(\sqrt{q_1^*}, \dots, \sqrt{q_t^*}) \subset L$. Поле $K(\sqrt{q_1^*}, \dots, \sqrt{q_t^*})$ зависит только от поля K (которое определяет d , но не f) и называется полем родов (*genus field*) для K . Далее будем обозначать $K(\sqrt{q_1^*}, \dots, \sqrt{q_t^*}) = K_G$.

3. Кольцо целых поля родов

Итак, введены в рассмотрение поля $K = \mathbb{Q}(\sqrt{d})$ и $K_G = K(\sqrt{q_1^*}, \dots, \sqrt{q_t^*})$. Далее понадобится знание кольца целых в K_G .

Пусть q_i^* взяты из условия леммы 2, тогда выполнен один из следующих трёх случаев:

- 1) все $|q_i|$ — нечётные простые;
- 2) $q_t^* = \pm 8$;
- 3) $q_t^* = -4$.

Построим в каждом из этих случаев фундаментальный базис поля K_G ; поскольку $d = q_1^* \dots q_t^*$, то $\sqrt{d} \in \mathbb{Q}(\sqrt{q_1^*}, \dots, \sqrt{q_t^*})$ и, следовательно, $K_G = \mathbb{Q}(\sqrt{q_1^*}, \dots, \sqrt{q_t^*})$.

Лемма 3. Пусть M — числовое поле, $p \in \mathbb{Z}$ — такое простое число, что идеал $p\mathbb{Z}$ неразветвлён в M . Пусть также $c \in M$ такое, что $pc^2 \in \mathcal{O}_M$. Тогда $c \in \mathcal{O}_M$.

Доказательство. Предположим, что $c \notin \mathcal{O}_M$. Тогда идеал $c\mathcal{O}_M$ дробный, и его разложение на простые имеет вид $\mathfrak{q}_1^{s_1} \dots \mathfrak{q}_m^{s_m}$, где все \mathfrak{q}_i попарно различны и $s_1 < 0$. Но степень \mathfrak{q}_1 в разложении идеала $p\mathcal{O}_M$ не превосходит 1 в силу неразветвлённости $p\mathbb{Z}$, а степень \mathfrak{q}_1 в разложении идеала $c^2\mathcal{O}_M$ не больше -2 , следовательно, степень идеала \mathfrak{q}_1 в разложении $pc^2\mathcal{O}_M$ отрицательна, что противоречит тому, что $pc^2 \in \mathcal{O}_M$. ■

Теорема 6. Пусть $\tilde{q}_1, \dots, \tilde{q}_r$ — такие попарно различные целые числа, что $|\tilde{q}_i|$ — нечётные простые и $\tilde{q}_i \equiv 1 \pmod{4}$. Обозначим $\alpha_i = (1 + \sqrt{\tilde{q}_i})/2$ и $\tilde{\alpha}_i = (1 - \sqrt{\tilde{q}_i})/2$. Тогда

- 1) числа $\alpha_1^{s_1} \dots \alpha_r^{s_r}$, когда набор (s_i) пробегает всевозможные наборы из $\{0, 1\}^r$, образуют фундаментальный базис поля $\mathbb{Q}(\sqrt{\tilde{q}_1}, \dots, \sqrt{\tilde{q}_r})$;
- 2) числа $\tilde{\alpha}_1^{s_1} \alpha_1^{1-s_1} \dots \tilde{\alpha}_r^{s_r} \alpha_r^{1-s_r}$, когда набор (s_i) пробегает всевозможные наборы из $\{0, 1\}^r$, образуют фундаментальный базис поля $\mathbb{Q}(\sqrt{\tilde{q}_1}, \dots, \sqrt{\tilde{q}_r})$.

Доказательство. Доказательство проведём индукцией по r . При $r = 0$ утверждение тривиально. Предположим, что утверждение доказано для полей $M_i = \mathbb{Q}(\sqrt{\tilde{q}_1}, \dots, \sqrt{\tilde{q}_i})$ при $i = 1, \dots, r-1$.

Лемма 4. Пусть $p \in \mathbb{Z}$ — простое число, не делящее ни одно из чисел $\tilde{q}_1, \dots, \tilde{q}_{r-1}$. Тогда идеал $p\mathbb{Z}$ неразветвлён в M_{r-1} .

Доказательство. Достаточно проверить, что для всех $1 \leq i \leq r-1$ никакой из идеалов поля M_{i-1} , делящих $p\mathcal{O}_{M_{i-1}}$, неразветвлён в $M_i = M_{i-1}(\sqrt{\tilde{q}_i})$.

Пусть \mathfrak{p} — простой идеал кольца целых поля M_{i-1} , такой, что $\mathfrak{p} \cap \mathbb{Z} = p\mathbb{Z}$. Расширение $M_{i-1} \subset M_i$ порождается элементом α_i ; из индуктивного предположения следует, что $(1, \alpha_i)$ образуют базис $\mathcal{O}_{M_i}/\mathcal{O}_{M_{i-1}}$. Единственный нетривиальный автоморфизм M_i над M_{i-1} переводит этот базис в $(1, \tilde{\alpha}_i)$. В соответствии с [20, Предложения III.8 и III.14] для доказательства неразветвлённости \mathfrak{p} достаточно проверить, что \mathfrak{p} не делит $\det \begin{pmatrix} 1 & \alpha_i \\ 1 & \tilde{\alpha}_i \end{pmatrix}^2 = (\alpha_i - \tilde{\alpha}_i)^2 = \tilde{q}_i$. Поскольку p не делит \tilde{q}_i , это свойство выполнено. ■

Применим лемму 4 к $p = |\tilde{q}_r|$. Разложение идеала $\tilde{q}_r\mathcal{O}_{M_{r-1}}$ на простые идеалы не содержит квадратов. В частности, $\sqrt{\tilde{q}_r} \notin M_{r-1}$, поскольку иначе было бы $(\tilde{q}_r\mathcal{O}_{M_{r-1}}) =$

$= (\sqrt{\tilde{q}_r} \mathcal{O}_{M_{r-1}})^2$. Следовательно, $(1, \alpha_r)$ образуют базис M_r над M_{r-1} . Пусть $a + b\alpha_r$ — целое алгебраическое число и $a, b \in M_{r-1}$. Число $a + b(1 - \alpha_r)$ — сопряжённое к $a + b\alpha_r$ и потому тоже является целым алгебраическим. Следовательно, их сумма $x = 2a + b$ и произведение $y = a^2 + ab + b^2(1 - \tilde{q}_r)/4$ также являются целыми алгебраическими и, следовательно, лежат в $\mathcal{O}_{M_{r-1}}$. Далее, $x^2 - 4y = \tilde{q}_r b^2 \in \mathcal{O}_{M_{r-1}}$. Из леммы 3 следует, что $b \in \mathcal{O}_{M_{r-1}}$. Следовательно, $2a \in \mathcal{O}_{M_{r-1}}$, $a^2 + ab \in \mathcal{O}_{M_{r-1}}$, $2a^2 = 2(a^2 + ab) - 2a \cdot b \in \mathcal{O}_{M_{r-1}}$. Применяя леммы 4 и 3 к $p = 2$, получаем $a \in \mathcal{O}_{M_{r-1}}$. Таким образом, если $a + b\alpha_r$ — целое алгебраическое и $a, b \in M_{r-1}$, то $a, b \in \mathcal{O}_{M_{r-1}}$; обратное утверждение очевидно, следовательно, $(1, \alpha_r)$ образуют базис кольца целых M_r над $\mathcal{O}_{M_{r-1}}$, что доказывает индуктивный переход для набора $\alpha_1^{s_1} \dots \alpha_r^{s_r}$. Для доказательства второго утверждения теоремы достаточно заметить, что $(1 - \alpha_r, \alpha_r) = ((1 - \sqrt{\tilde{q}_r})/2, (1 + \sqrt{\tilde{q}_r})/2)$ также образуют базис кольца целых M_r над $\mathcal{O}_{M_{r-1}}$. ■

Теорема 7. Пусть $\tilde{q}_1, \dots, \tilde{q}_{r-1}$ такие же, как в теореме 6, и $\tilde{q}_r = \pm 8$. Обозначим $\alpha_r = \sqrt{\tilde{q}_r}/4$. Тогда следующие множества образуют фундаментальные базисы поля $\mathbb{Q}(\sqrt{\tilde{q}_1}, \dots, \sqrt{\tilde{q}_r})$:

- 1) числа $\alpha_1^{s_1} \dots \alpha_r^{s_r}$, когда набор (s_i) пробегает всевозможные наборы из $\{0, 1\}^r$;
- 2) числа $\tilde{\alpha}_1^{s_1} \alpha_1^{1-s_1} \dots \tilde{\alpha}_{r-1}^{s_{r-1}} \alpha_{r-1}^{1-s_{r-1}} \alpha_r^{s_r}$, когда набор (s_i) пробегает всевозможные наборы из $\{0, 1\}^r$.

Доказательство. Положим $M = \mathbb{Q}(\sqrt{\tilde{q}_1}, \dots, \sqrt{\tilde{q}_{r-1}})$. Применим лемму 4 к $p = 2$ и теорему 6. Идеал $2\mathbb{Z}$ неразветвлён в M , следовательно, как было показано ранее, $\sqrt{\tilde{q}_r} \notin M$ и $(1, \sqrt{\tilde{q}_r})$ образуют базис $M(\sqrt{\tilde{q}_r})$ над M .

Пусть $a + b\alpha_r$ — целое алгебраическое число и $a, b \in M$. Число $a - b\alpha_r$ — сопряжённое к $a + b\alpha_r$ и потому тоже является целым алгебраическим. Следовательно, их сумма $2a$ и произведение $a^2 \mp 2b^2$ также являются целыми алгебраическими и, следовательно, лежат в \mathcal{O}_M . Далее, $(2a)^2 - 4(a^2 \mp 2b^2) = \pm 2(2b)^2 \in \mathcal{O}_M$, по лемме 3 отсюда следует, что $2b \in \mathcal{O}_M$. Теперь $2(a^2 \mp 2b^2) \pm (2b)^2 = 2a^2 \in \mathcal{O}_M$ и, повторно применяя лемму 3, находим $a \in \mathcal{O}_M$. Наконец, $a^2 - (a^2 \mp 2b^2) = \pm 2b^2 \in \mathcal{O}_M$, и третье применение леммы 3 даёт $b \in \mathcal{O}_M$. Таким образом, $(1, \alpha_r)$ образуют базис кольца целых поля $M(\sqrt{\tilde{q}_r}^*)$ над \mathcal{O}_M , и применение теоремы 6 завершает доказательство. ■

Теорема 8. Пусть $\tilde{q}_1, \dots, \tilde{q}_{r-1}$ такие же, как в теореме 6, и $\tilde{q}_r = -4$. Обозначим $\alpha_r = \sqrt{\tilde{q}_r}/4 = i$. Тогда следующие множества образуют фундаментальные базисы поля $\mathbb{Q}(\sqrt{\tilde{q}_1}, \dots, \sqrt{\tilde{q}_r})$:

- 1) числа $\alpha_1^{s_1} \dots \alpha_r^{s_r}$, когда набор (s_i) пробегает всевозможные наборы из $\{0, 1\}^r$;
- 2) числа $\tilde{\alpha}_1^{s_1} \alpha_1^{1-s_1} \dots \tilde{\alpha}_{r-1}^{s_{r-1}} \alpha_{r-1}^{1-s_{r-1}} \alpha_r^{s_r}$, когда набор (s_i) пробегает всевозможные наборы из $\{0, 1\}^r$.

Доказательство. Положим $M = \mathbb{Q}(\sqrt{\tilde{q}_1}, \dots, \sqrt{\tilde{q}_{r-1}})$. Равенство $2 = -i(1 + i)^2$ показывает, что идеал $2\mathbb{Z}$ разветвлён в любом поле, содержащем i . Согласно лемме 4 и теореме 6, $2\mathbb{Z}$ неразветвлён в M . Следовательно, $i \notin M$.

Пусть $a + bi$ — целое алгебраическое число и $a, b \in M$. Число $a - bi$ — сопряжённое к $a + bi$ и потому тоже является целым алгебраическим. Следовательно, их сумма $2a$ и произведение $a^2 + b^2$ также являются целыми алгебраическими и, следовательно, лежат в \mathcal{O}_M . Далее, $2(a^2 + b^2) + 2a \cdot 2b = 2(a + b)^2 \in \mathcal{O}_M$ и по леммам 4 и 3, применённым к $p = 2$, и теореме 6 $a + b \in \mathcal{O}_M$. Теперь $2a - (a + b) = a - b \in \mathcal{O}_M$, $(a + b)(a - b) = a^2 - b^2 \in \mathcal{O}_M$, $2a^2 \in \mathcal{O}_M$, $2b^2 \in \mathcal{O}_M$; снова применяя леммы 4 и 3 и теорему 6, получаем $a, b \in \mathcal{O}_M$. Таким образом, $(1, \alpha_t)$ образуют базис кольца целых поля $M(\sqrt{\tilde{q}_r})$ над \mathcal{O}_M , и применение теоремы 6 завершает доказательство. ■

Поскольку в каждом случае из теорем следует, что $[K_G : \mathbb{Q}] = 2^t$, то $\sqrt{q_j^*} \notin \mathbb{Q}(\dots, \sqrt{q_{j-1}^*}, \sqrt{q_{j+1}^*}, \dots)$ для любого $1 \leq j \leq t$. Следовательно, в $\text{Gal}(K_G/\mathbb{Q})$ есть t элементов τ_j , действующих следующим образом:

$$\tau_j \left(\sqrt{q_j^*} \right) = -\sqrt{q_j^*}, \quad \tau_j \left(\sqrt{q_i^*} \right) = \sqrt{q_i^*} \text{ при } i \neq j. \quad (14)$$

Положим $\tau'_\mu = \tau_1^{\mu_1} \dots \tau_t^{\mu_t} \in \text{Gal}(K_G/\mathbb{Q})$ при $\mu \in \{0, 1\}^t$; сравнивая действие τ'_μ на $\sqrt{q_i^*}$, легко видеть, что все τ'_μ различны. Получаем $2^t = |\text{Gal}(K_G/\mathbb{Q})|$ различных элементов $\text{Gal}(K_G/\mathbb{Q})$, которые, таким образом, исчерпывают эту группу.

Предыдущие теоремы задают \mathbb{Z} -базис \mathcal{O}_{K_G} . Нас также будут интересовать пересечение \mathcal{O}_{K_G} с \mathbb{R} (являющееся, очевидно, кольцом целых в поле $K_G \cap \mathbb{R}$) и пересечение \mathcal{O}_{K_G} с $i\mathbb{R}$ (являющееся, очевидно, \mathbb{Z} -модулем). Среди q_i^* есть отрицательные числа. Обозначив через u число положительных среди q_i^* , $0 \leq u < t$, можно без ограничения общности считать, что $q_1^* > 0, \dots, q_u^* > 0, q_{u+1}^* < 0, \dots, q_t^* < 0$.

Комплексное сопряжение действует на $\sqrt{q_i^*}$ так же, как и композиция $\tau_{u+1} \dots \tau_t$. Поскольку $K_G \cap \mathbb{R}$ — это максимальное подполе K_G , инвариантное относительно комплексного сопряжения, то $\text{Gal}((K_G \cap \mathbb{R})/\mathbb{Q})$ изоморфна фактор-группе $\text{Gal}(K_G/\mathbb{Q})$ по подгруппе, порождённой комплексным сопряжением. Выбирая в качестве представителя смежного класса элемент с $\mu_t = 0$, получаем, что $\text{Gal}((K_G \cap \mathbb{R})/\mathbb{Q})$ состоит из автоморфизмов

$$\tau_\lambda = \tau_{\lambda_1, \dots, \lambda_{t-1}} = \tau'_{\lambda_1, \dots, \lambda_{t-1}, 0} = \tau_1^{\lambda_1} \dots \tau_{t-1}^{\lambda_{t-1}} \quad (15)$$

при $\lambda \in \{0, 1\}^{t-1}$, различных при разных λ .

В дальнейшем из двух значений \sqrt{d} будем выбирать то, которое соответствует произведению $\sqrt{q_1^*} \dots \sqrt{q_t^*}$, где значения отдельных корней такие же, как при вычислении α_i и $\tilde{\alpha}_i$.

Теорема 9. Пусть q_1^*, \dots, q_t^* такие же, как в лемме 2, нечётные, занумерованные так, что $q_i^* > 0$ при $1 \leq i \leq u$ и $q_i^* < 0$ при $u < i \leq t$, где u — некоторое число от 0 до $t-1$ включительно; $K_G = \mathbb{Q}(\sqrt{q_1^*}, \dots, \sqrt{q_t^*})$; τ_λ заданы формулой (15). Тогда

1) числа

$$\begin{aligned} \beta_{s_1, \dots, s_{t-1}} &= \beta_{s_1, \dots, s_{t-1}}(q_1^*, \dots, q_t^*) = \\ &= \left(\prod_{i=1}^u \tilde{\alpha}_i^{s_i} \alpha_i^{1-s_i} \right) \left(\left(\prod_{i=u+1}^{t-1} \tilde{\alpha}_i^{s_i} \alpha_i^{1-s_i} \right) \alpha_t + \left(\prod_{i=u+1}^{t-1} \tilde{\alpha}_i^{1-s_i} \alpha_i^{s_i} \right) \tilde{\alpha}_t \right), \end{aligned}$$

когда набор (s_i) пробегает всевозможные наборы из $\{0, 1\}^{t-1}$, образуют фундаментальный базис поля $K_G \cap \mathbb{R}$;

2) числа

$$\begin{aligned} \beta_{s_1, \dots, s_{t-1}}^* &= \beta_{s_1, \dots, s_{t-1}}^*(q_1^*, \dots, q_t^*) = \\ &= \left(\prod_{i=1}^u (-\tilde{\alpha}_i)^{s_i} \alpha_i^{1-s_i} \right) \left(\left(\prod_{i=u+1}^{t-1} (-\tilde{\alpha}_i)^{s_i} \alpha_i^{1-s_i} \right) \alpha_t - \left(\prod_{i=u+1}^{t-1} \tilde{\alpha}_i^{1-s_i} (-\alpha_i)^{s_i} \right) \tilde{\alpha}_t \right), \end{aligned}$$

когда набор (s_i) пробегает всевозможные наборы из $\{0, 1\}^{t-1}$, образуют базис \mathbb{Z} -модуля $\mathcal{O}_{K_G} \cap i\mathbb{R}$;

3) для любых $\eta, \nu \in \{0, 1\}^{t-1}$ справедливо равенство

$$\sum_{\mu \in \{0, 1\}^{t-1}} (-1)^{\mu_1 + \dots + \mu_{t-1}} \tau_\mu \left(\beta_{\eta_1, \dots, \eta_{t-1}} \beta_{\nu_1, \dots, \nu_{t-1}}^* \right) = \begin{cases} \sqrt{d}, & \text{если } \eta = \nu, \\ 0, & \text{иначе.} \end{cases}$$

Доказательство. Обозначим элемент базиса из п. 2 теоремы 6, соответствующий набору (s_1, \dots, s_t) , через β'_{s_1, \dots, s_t} .

Число из \mathcal{O}_{K_G} попадает в пересечение $K_G \cap \mathbb{R}$ тогда и только тогда, когда оно инвариантно относительно комплексного сопряжения. Легко видеть, что комплексное сопряжение переводит β'_{s_1, \dots, s_t} в $\beta'_{s_1, \dots, s_u, 1-s_{u+1}, \dots, 1-s_t}$. Таким образом, для инвариантности \mathbb{Z} -линейной комбинации β'_{s_1, \dots, s_t} необходимо и достаточно, чтобы для каждого набора (s_i) коэффициенты при β'_{s_1, \dots, s_t} и $\beta'_{s_1, \dots, s_u, 1-s_{u+1}, \dots, 1-s_t}$ совпадали. Теперь из теоремы 6 следует, что $\{\beta'_{s_1, \dots, s_{t-1}, 0} + \beta'_{s_1, \dots, s_u, 1-s_{u+1}, \dots, 1-s_{t-1}, 1}\}$ образуют требуемый фундаментальный базис. Из определения β' легко видеть, что эта сумма равна $\beta_{s_1, \dots, s_{t-1}}$, что доказывает первое утверждение теоремы.

Число из \mathcal{O}_{K_G} попадает в пересечение $K_G \cap i\mathbb{R}$ тогда и только тогда, когда оно меняет знак под действием комплексного сопряжения. Аналогично предыдущему получаем, что $\{\beta'_{s_1, \dots, s_{t-1}, 0} - \beta'_{s_1, \dots, s_u, 1-s_{u+1}, \dots, 1-s_{t-1}, 1}\}$ образуют требуемый \mathbb{Z} -базис. Из определения β' легко видеть, что эта разность равна $\pm \beta_{s_1, \dots, s_{t-1}}^*$, что доказывает второе утверждение теоремы.

Наконец, третье утверждение проверяется прямой выкладкой. А именно, легко видеть, что

$$\begin{aligned} & \tau_\mu(\beta_{\eta_1, \dots, \eta_{t-1}}) = \\ & = \left(\prod_{i=1}^u \tilde{\alpha}_i^{\mu_i \oplus \eta_i} \alpha_i^{1-(\mu_i \oplus \eta_i)} \right) \left(\left(\prod_{i=u+1}^{t-1} \tilde{\alpha}_i^{\mu_i \oplus \eta_i} \alpha_i^{1-(\mu_i \oplus \eta_i)} \right) \alpha_t + \left(\prod_{i=u+1}^{t-1} \tilde{\alpha}_i^{1-(\mu_i \oplus \eta_i)} \alpha_i^{\mu_i \oplus \eta_i} \right) \tilde{\alpha}_t \right), \\ & \tau_\mu(\beta_{\nu_1, \dots, \nu_{t-1}}^*) = \left(\prod_{i=1}^u (-1)^{\nu_i} \tilde{\alpha}_i^{\mu_i \oplus \nu_i} \alpha_i^{1-(\mu_i \oplus \nu_i)} \right) \times \\ & \times \left(\left(\prod_{i=u+1}^{t-1} (-1)^{\nu_i} \tilde{\alpha}_i^{\mu_i \oplus \nu_i} \alpha_i^{1-(\mu_i \oplus \nu_i)} \right) \alpha_t - \left(\prod_{i=u+1}^{t-1} (-1)^{\nu_i} \tilde{\alpha}_i^{1-(\mu_i \oplus \nu_i)} \alpha_i^{\mu_i \oplus \nu_i} \right) \tilde{\alpha}_t \right). \end{aligned}$$

Подставим эти формулы в произведение $\tau_\mu(\beta_{\eta_1, \dots, \eta_{t-1}}) \tau_\mu(\beta_{\nu_1, \dots, \nu_{t-1}}^*)$, получим формулу вида $(a+b)(c-d)$, в которой раскроем скобки, получив четыре слагаемых $ac+bc-ad-bd$. Под δ_{ij} будем понимать символ Кронекера: $\delta_{ii} = 1$, $\delta_{ij} = 0$ при $i \neq j$. Заметим, что

$$\begin{aligned} & \sum_{\mu_i=0}^1 (-1)^{\mu_i} (-1)^{\nu_i} \tilde{\alpha}_i^{(\mu_i \oplus \eta_i) + (\mu_i \oplus \nu_i)} \alpha_i^{1-(\mu_i \oplus \eta_i) + 1-(\mu_i \oplus \nu_i)} = \\ & = (-1)^{\nu_i} \left(\tilde{\alpha}_i^{\eta_i + \nu_i} \alpha_i^{2-(\eta_i + \nu_i)} - \tilde{\alpha}_i^{2-(\eta_i + \nu_i)} \alpha_i^{\eta_i + \nu_i} \right) = \delta_{\eta_i \nu_i} (\alpha_i^2 - \tilde{\alpha}_i^2) = \delta_{\eta_i \nu_i} \sqrt{q_i^*}, \\ & \sum_{\mu_i=0}^1 (-1)^{\mu_i} (-1)^{\nu_i} \tilde{\alpha}_i^{1-(\mu_i \oplus \eta_i) + (\mu_i \oplus \nu_i)} \alpha_i^{(\mu_i \oplus \eta_i) + 1-(\mu_i \oplus \nu_i)} = \\ & = (-1)^{\nu_i} \left(\tilde{\alpha}_i^{1-\eta_i + \nu_i} \alpha_i^{1+\eta_i - \nu_i} - \tilde{\alpha}_i^{1+\eta_i - \nu_i} \alpha_i^{1-\eta_i + \nu_i} \right) = \delta_{\eta_i + \nu_i, 1} (\alpha_i^2 - \tilde{\alpha}_i^2) = \delta_{\eta_i + \nu_i, 1} \sqrt{q_i^*} \end{aligned}$$

и перестановка α_i и $\tilde{\alpha}_i$ даёт ещё два рассматриваемых произведения, которые совпадают с предыдущими, умноженными на (-1) .

Таким образом,

$$\begin{aligned} & \sum_{\mu \in \{0,1\}^{t-1}} (-1)^{\mu_1 + \dots + \mu_{t-1}} \tau_\mu(\beta_{\eta_1, \dots, \eta_{t-1}} \beta_{\nu_1, \dots, \nu_{t-1}}^*) = \\ & = \left(\prod_{i=1}^u \delta_{\eta_i \nu_i} \sqrt{q_i^*} \right) \left(\alpha_t^2 \prod_{i=u+1}^{t-1} \delta_{\eta_i \nu_i} \sqrt{q_i^*} + \tilde{\alpha}_t \alpha_t \prod_{i=u+1}^{t-1} \delta_{\eta_i + \nu_i, 1} \sqrt{q_i^*} - \right. \\ & \left. - \alpha_t \tilde{\alpha}_t \prod_{i=u+1}^{t-1} (-\delta_{\eta_i + \nu_i, 1} \sqrt{q_i^*}) - \tilde{\alpha}_t^2 \prod_{i=u+1}^{t-1} (-\delta_{\eta_i \nu_i} \sqrt{q_i^*}) \right). \end{aligned}$$

Знак произведения $q_1^* \cdots q_t^*$ определяется чётностью количества отрицательных множителей, которых ровно $t - u$. Таким образом, из условия $q_1^* \cdots q_t^* < 0$ следует, что

$$t - u \text{ нечётно и } \prod_{i=u+1}^{t-1} (-1) = (-1)^{t-u-1} = 1;$$

$$\begin{aligned} & \sum_{\mu \in \{0,1\}^{t-1}} (-1)^{\mu_1 + \dots + \mu_{t-1}} \tau_\mu (\beta_{\eta_1, \dots, \eta_{t-1}} \beta_{\nu_1, \dots, \nu_{t-1}}^*) = \\ & = \left(\prod_{i=1}^u \delta_{\eta_i \nu_i} \sqrt{q_i^*} \right) (\alpha_t^2 - \tilde{\alpha}_t^2) \left(\prod_{i=u+1}^{t-1} \delta_{\eta_i \nu_i} \sqrt{q_i^*} \right) = \sqrt{q_t^*} \prod_{i=1}^{t-1} \delta_{\eta_i \nu_i} \sqrt{q_i^*}. \end{aligned}$$

Теорема доказана. ■

Теорема 10. Пусть q_2^*, \dots, q_t^* такие же, как в теореме 9, и $q_1^* = 8$. Пусть q_i^* занумерованы так, что $q_i^* > 0$ при $1 \leq i \leq u$ и $q_i^* < 0$ при $u < i \leq t$, где u — некоторое число от 1 до $t - 1$ включительно; $K_G = \mathbb{Q}(\sqrt{q_1^*}, \dots, \sqrt{q_t^*})$; τ_λ заданы формулой (15). Тогда

1) числа

$$\begin{aligned} \beta_{s_1, \dots, s_{t-1}} &= \beta_{s_1, \dots, s_{t-1}}(q_1^*, \dots, q_t^*) = \sqrt{2}^{s_1} \left(\prod_{i=2}^u \tilde{\alpha}_i^{s_i} \alpha_i^{1-s_i} \right) \times \\ & \times \left(\left(\prod_{i=u+1}^{t-1} \tilde{\alpha}_i^{s_i} \alpha_i^{1-s_i} \right) \alpha_t + \left(\prod_{i=u+1}^{t-1} \tilde{\alpha}_i^{1-s_i} \alpha_i^{s_i} \right) \tilde{\alpha}_t \right), \end{aligned}$$

когда набор (s_i) пробегает всевозможные наборы из $\{0, 1\}^{t-1}$, образуют фундаментальный базис поля $K_G \cap \mathbb{R}$;

2) числа

$$\begin{aligned} \beta_{s_1, \dots, s_{t-1}}^* &= \beta_{s_1, \dots, s_{t-1}}^*(q_1^*, \dots, q_t^*) = \sqrt{2}^{1-s_1} \left(\prod_{i=2}^u (-\tilde{\alpha}_i)^{s_i} \alpha_i^{1-s_i} \right) \times \\ & \times \left(\left(\prod_{i=u+1}^{t-1} (-\tilde{\alpha}_i)^{s_i} \alpha_i^{1-s_i} \right) \alpha_t - \left(\prod_{i=u+1}^{t-1} \tilde{\alpha}_i^{1-s_i} (-\alpha_i)^{s_i} \right) \tilde{\alpha}_t \right), \end{aligned}$$

когда набор (s_i) пробегает всевозможные наборы из $\{0, 1\}^{t-1}$, образуют базис \mathbb{Z} -модуля $\mathcal{O}_{K_G} \cap i\mathbb{R}$;

3) для любых $\nu, \eta \in \{0, 1\}^{t-1}$ справедливо равенство

$$\sum_{\mu \in \{0,1\}^{t-1}} (-1)^{\mu_1 + \dots + \mu_{t-1}} \tau_\mu (\beta_{\eta_1, \dots, \eta_{t-1}} \beta_{\nu_1, \dots, \nu_{t-1}}^*) = \begin{cases} \sqrt{d}, & \text{если } \eta = \nu, \\ 0, & \text{иначе.} \end{cases}$$

Доказательство аналогично доказательству теоремы 9. При вычислении выражения из третьего утверждения теоремы по сравнению с теоремой 9 возникает дополнительный множитель

$$\sum_{\mu_1=0}^1 (-1)^{\mu_1} ((-1)^{\mu_1} \sqrt{2})^{\eta_1} ((-1)^{\mu_1} \sqrt{2})^{1-\nu_1} = \sqrt{2}^{1+\eta_1-\nu_1} (1 + (-1)^{\eta_1+\nu_1}) = 2\sqrt{2} \delta_{\eta_1 \nu_1}. \quad \blacksquare$$

Теорема 11. Пусть q_1^*, \dots, q_{t-1}^* такие же, как в теореме 9, и $q_t^* \in \{-4, -8\}$. Пусть q_i^* занумерованы так, что $q_i^* > 0$ при $1 \leq i \leq u$ и $q_i^* < 0$ при $u < i \leq t$, где u — некоторое число от 0 до $t - 2$ включительно; $K_G = \mathbb{Q}(\sqrt{q_1^*}, \dots, \sqrt{q_t^*})$; τ_λ заданы формулой (15). Тогда

1) числа

$$\beta_{s_1, \dots, s_{t-1}} = \beta_{s_1, \dots, s_{t-1}}(\sqrt{q_1^*}, \dots, \sqrt{q_t^*}) = \left(\prod_{i=1}^u \tilde{\alpha}_i^{s_i} \alpha_i^{1-s_i} \right) \times \\ \times \left(\left(\prod_{i=u+1}^{t-2} \tilde{\alpha}_i^{s_i} \alpha_i^{1-s_i} \right) \alpha_{t-1} \alpha_t^{s_{t-1}} + \left(\prod_{i=u+1}^{t-2} \tilde{\alpha}_i^{1-s_i} \alpha_i^{s_i} \right) \tilde{\alpha}_{t-1} (-\alpha_t)^{s_{t-1}} \right),$$

когда набор (s_i) пробегает всевозможные наборы из $\{0, 1\}^{t-1}$, образуют фундаментальный базис поля $K_G \cap \mathbb{R}$;

2) числа

$$\beta_{s_1, \dots, s_{t-1}}^* = \beta_{s_1, \dots, s_{t-1}}^*(\sqrt{q_1^*}, \dots, \sqrt{q_t^*}) = \left(\prod_{i=1}^u (-\tilde{\alpha}_i)^{s_i} \alpha_i^{1-s_i} \right) \times \\ \times \left(\left(\prod_{i=u+1}^{t-2} (-\tilde{\alpha}_i)^{s_i} \alpha_i^{1-s_i} \right) \alpha_{t-1} \alpha_t^{1-s_{t-1}} - \left(\prod_{i=u+1}^{t-2} \tilde{\alpha}_i^{1-s_i} (-\alpha_i)^{s_i} \right) \tilde{\alpha}_{t-1} (-\alpha_t)^{1-s_{t-1}} \right),$$

когда набор (s_i) пробегает всевозможные наборы из $\{0, 1\}^{t-1}$, образуют базис \mathbb{Z} -модуля $\mathcal{O}_{K_G} \cap i\mathbb{R}$;

3) для любых $\eta, \nu \in \{0, 1\}^{t-1}$ справедливо равенство

$$\sum_{\mu \in \{0, 1\}^{t-1}} (-1)^{\mu_1 + \dots + \mu_{t-1}} \tau_\mu (\beta_{\eta_1, \dots, \eta_{t-1}} \beta_{\nu_1, \dots, \nu_{t-1}}^*) = \begin{cases} \sqrt{d}, & \text{если } \eta = \nu, \\ 0, & \text{иначе.} \end{cases}$$

Доказательство. Обозначим элемент базиса из п. 2 теоремы 8, соответствующий набору (s_1, \dots, s_t) , через β'_{s_1, \dots, s_t} .

Число из кольца целых поля K_G попадает в пересечение этого поля с \mathbb{R} тогда и только тогда, когда оно инвариантно относительно комплексного сопряжения. Легко видеть, что комплексное сопряжение переводит β'_{s_1, \dots, s_t} в $(-1)^{s_t} \beta'_{s_1, \dots, s_u, 1-s_{u+1}, \dots, 1-s_{t-1}, s_t}$. Следовательно, для инвариантности линейной комбинации β'_{s_1, \dots, s_t} с коэффициентами из \mathbb{Z} необходимо и достаточно, чтобы для каждого набора (s_i) коэффициенты при β'_{s_1, \dots, s_t} и $\beta'_{s_1, \dots, s_u, 1-s_{u+1}, \dots, 1-s_{t-1}, s_t}$ отличались друг от друга умножением на $(-1)^{s_t}$. Теперь из теоремы 8 следует, что $\{\beta'_{s_1, \dots, s_{t-2}, 0, s_t} + (-1)^{s_t} \beta'_{s_1, \dots, s_u, 1-s_{u+1}, \dots, 1-s_{t-2}, 1, s_t}\}$ образуют требуемый фундаментальный базис. Из определения β' легко видеть, что эта сумма равна $\beta_{s_1, \dots, s_{t-2}, s_t}$, что доказывает первое утверждение теоремы.

Доказательство второго утверждения аналогично первому.

Третье утверждение доказывается явной выкладкой. Аналогично доказательству теоремы 9 находим

$$\sum_{\mu \in \{0, 1\}^{t-1}} (-1)^{\mu_1 + \dots + \mu_{t-1}} \tau_\mu (\beta_{\eta_1, \dots, \eta_{t-1}} \beta_{\nu_1, \dots, \nu_{t-1}}^*) = \\ = \left(\prod_{i=1}^u \delta_{\eta_i \nu_i} \sqrt{q_i^*} \right) \alpha_t^{\eta_{t-1} + 1 - \nu_{t-1}} \left(\sqrt{q_{t-1}^*} \prod_{i=u+1}^{t-2} \delta_{\eta_i \nu_i} \sqrt{q_i^*} + \right. \\ \left. + (-1)^{\nu_{t-1} + \eta_{t-1}} (-\sqrt{q_{t-1}^*}) \prod_{i=u+1}^{t-2} (-\delta_{\eta_i \nu_i} \sqrt{q_i^*}) \right).$$

Поскольку $q_1^* \dots q_t^* < 0$, то количество отрицательных среди q_i^* , то есть $t - u$, нечётно и, следовательно, $\prod_{i=u+1}^{t-2} (-1) = (-1)^{t-u-2} = -1$. Тогда

$$\begin{aligned} & \sum_{\mu \in \{0,1\}^{t-1}} (-1)^{\mu_1 + \dots + \mu_{t-1}} \tau_\mu (\beta_{\eta_1, \dots, \eta_{t-1}} \beta_{\nu_1, \dots, \nu_{t-1}}^*) = \\ & = \left(\prod_{i=1}^{t-2} \delta_{\eta_i \nu_i} \sqrt{q_i^*} \right) (1 + (-1)^{\nu_{t-1} + \eta_{t-1}}) \alpha_t^{1 + \eta_{t-1} - \nu_{t-1}} = \delta_{\eta_{t-1} \nu_{t-1}} \left(\prod_{i=1}^{t-2} \delta_{\eta_i \nu_i} \sqrt{q_i^*} \right) 2\alpha_t. \end{aligned}$$

Теорема доказана. ■

Теорема 12. Пусть q_1^*, \dots, q_{t-1}^* нечётные положительные, $q_t^* = -4$ или $q_t^* = -8$. Тогда

- 1) числа $\beta_{s_1, \dots, s_{t-1}} = \beta_{s_1, \dots, s_{t-1}}(q_1^*, \dots, q_t^*) = \prod_{i=1}^{t-1} \tilde{\alpha}_i^{s_i} \alpha_i^{1-s_i}$, когда набор (s_i) пробегает всевозможные наборы из $\{0, 1\}^{t-1}$, образуют фундаментальный базис поля $K_G \cap \mathbb{R}$;
- 2) числа $\beta_{s_1, \dots, s_{t-1}}^* = \beta_{s_1, \dots, s_{t-1}}^*(q_1^*, \dots, q_t^*) = \left(\prod_{i=1}^{t-1} (-\tilde{\alpha}_i)^{s_i} \alpha_i^{1-s_i} \right) \sqrt{q_t^*}$, когда набор (s_i) пробегает всевозможные наборы из $\{0, 1\}^{t-1}$, образуют базис \mathbb{Z} -модуля $\mathcal{O}_{K_G} \cap i\mathbb{R}$;
- 3) для любых $\eta, \nu \in \{0, 1\}^{t-1}$ справедливо равенство

$$\sum_{\mu \in \{0,1\}^{t-1}} (-1)^{\mu_1 + \dots + \mu_{t-1}} \tau_\mu (\beta_{\eta_1, \dots, \eta_{t-1}} \beta_{\nu_1, \dots, \nu_{t-1}}^*) = \begin{cases} \sqrt{d}, & \text{если } \eta = \nu, \\ 0, & \text{иначе.} \end{cases}$$

Доказательство. Очевидно, что в условиях теоремы $K_G = M(\sqrt{q_t^*})$, где $M \subset \mathbb{R}$. Следовательно, $K_G \cap \mathbb{R} = M$, $K_G \cap i\mathbb{R} = \sqrt{q_t^*} \cdot M$. Первые два утверждения следуют из теоремы 6, третье доказывается выкладкой, аналогичной доказательству теоремы 9. ■

Будем обозначать $\beta_\mu = \beta_{\mu_1, \dots, \mu_{t-1}}$ при $\mu \in \{0, 1\}^{t-1}$. Множество $\{\beta_\mu\}$ образует фундаментальный базис поля $K_G \cap \mathbb{R}$, которое в дальнейшем будем обозначать M .

Пусть z — произвольный элемент \mathcal{O}_{K_G} . Поскольку $z \in K_G$, то и $\bar{z} \in K_G$, так что $z + \bar{z} = 2 \operatorname{Re} z \in K_G \cap \mathbb{R}$ и $z - \bar{z} = 2i \operatorname{Im} z \in K_G \cap i\mathbb{R}$. Более того, z и, следовательно, \bar{z} являются целыми алгебраическими, а потому $2 \operatorname{Re} z$ и $2i \operatorname{Im} z$ также являются целыми алгебраическими. Таким образом, $2 \operatorname{Re} z = \sum_{\mu} b_\mu \beta_\mu$ и $2i \operatorname{Im} z = \sum_{\mu} b'_\mu \beta_\mu^*$. Здесь и далее, если явным образом не указано множество значений параметра, записанного греческой буквой, подразумевается $\{0, 1\}^{t-1}$. Опишем, как по приближённому значению таких сумм находить наборы чисел b_μ и b'_μ . Числа β_μ образуют базис вещественного поля M , базис β_μ^* чисто мнимый и становится базисом того же поля M после деления, например, на $\sqrt{q_t^*} \in K_G$, $q_t^* < 0$. Таким образом, можно найти точное выражение для z по приближённому значению, если уметь решать задачу восстановления целых коэффициентов разложения вещественного числа, заданного с некоторой точностью, по вещественному базису.

Дальнейший план изложения таков.

- Выделим в $\mathcal{O}_{K_G}[x]$ делитель многочлена $H_D[\theta, \alpha_*]$ степени $h/2^{t-1}$, который будем рассматривать вместо многочлена $H_D[\theta, \alpha_*]$. Это позволит уменьшить число вычисляемых коэффициентов и их величину. Этому посвящён п. 4.
- Оценим сверху все сопряжённые числа к коэффициентам рассматриваемого многочлена (п. 5).

- Основная идея для восстановления коэффициентов — использовать рациональные приближения к элементам базиса с одним и тем же знаменателем достаточной точности в зависимости от оценки из п. 5. Построению таких приближений для случая β_μ и β_μ^* посвящён п. 6.
- Далее в п. 7 опишем, как при помощи этих рациональных приближений точно определить коэффициенты рассматриваемого многочлена, вычисленные по определению с некоторой точностью.

4. Делитель многочлена $H_D[\theta, \alpha_*](x)$

Пусть $\tilde{\mathfrak{a}} \in \mathcal{H}_D$. Выберем форму (A, B, C) так, чтобы $\mathfrak{h}(A, B, C) = \tilde{\mathfrak{a}}$ и $\gcd(A, D) = 1$; это возможно, поскольку \mathfrak{h} зависит только от класса эквивалентности формы, а в каждом классе эквивалентных форм, согласно [5, Lemma 2.25, Lemma 2.3], есть представитель (A, B, C) с $\gcd(A, D) = 1$. Определим отображение $\varphi : \mathcal{H}_D \rightarrow \{\pm 1\}^t$ следующей формулой:

$$\varphi(\tilde{\mathfrak{a}}) = \left(\left(\frac{q_1^*}{A} \right), \dots, \left(\frac{q_t^*}{A} \right) \right).$$

Это определение корректно, поскольку отображение Артина зависит только от класса идеала в $I(\mathcal{O}, f)/P_{K, \mathbb{Z}}(f)$, из теоремы 5 следует, что $\left(\frac{q_i^*}{A} \right)$ не меняется при замене формы (A, B, C) на эквивалентную с сохранением условия $\gcd(A, D) = 1$.

Теорема 13. Образ отображения φ совпадает с группой $\{(\varepsilon_1, \dots, \varepsilon_t) \in \{\pm 1\}^t : \prod_i \varepsilon_i = 1\}$ с покомпонентным умножением. Отображение φ является гомоморфизмом групп. Поле инвариантов $L^{\Omega(\text{Ker } \varphi)} = \{x \in L : \tau(x) = x \text{ для всех } \tau \in \Omega(\text{Ker } \varphi)\}$ есть $K(\sqrt{q_1^*}, \dots, \sqrt{q_t^*})$.

Доказательство. То, что φ является гомоморфизмом, следует из п. 3 теоремы 5 и того, что отображение Артина является гомоморфизмом.

Проверим, что для любого элемента из образа φ произведение его компонент равно 1. Рассмотрим идеал \mathfrak{a} , определённый, как в теореме 5, для формы (A, B, C) . Тогда

$$\left(\frac{q_i^*}{A} \right) = \frac{1}{\sqrt{q_i^*}} \left(\frac{L/K}{\Omega_1(\mathfrak{a})} \right) (\sqrt{q_i^*})$$

и, перемножая эти равенства по всем i , с учётом леммы 2 находим

$$\left(\frac{q_1^*}{A} \right) \cdot \dots \cdot \left(\frac{q_t^*}{A} \right) = \frac{1}{\sqrt{d}} \left(\frac{L/K}{\Omega_1(\mathfrak{a})} \right) (\sqrt{d}).$$

Но $\sqrt{d} \in K$, а $\left(\frac{L/K}{\Omega_1(\mathfrak{a})} \right)$, будучи элементом $\text{Gal}(L/K)$, оставляет на месте элементы K .

Проверим, что подгруппа $\Omega(\text{Ker } \varphi)$ оставляет на месте элементы $\sqrt{q_i^*}$. Пусть $\varphi(\tilde{\mathfrak{a}}) = (1, \dots, 1)$, \mathfrak{a} — представитель $\tilde{\mathfrak{a}}$ из п. 2 теоремы 5. Тогда

$$\left(\frac{L/K}{\Omega_1(\mathfrak{a})} \right) (\sqrt{q_i^*}) = \sqrt{q_i^*},$$

то есть образ $\Omega_1(\mathfrak{a})$ при отображении Артина, равный $\Omega(\tilde{\mathfrak{a}})$ в силу коммутативности диаграммы (12), действует на $\sqrt{q_i^*}$ тривиально, что и требовалось доказать.

Таким образом,

$$\text{Im } \varphi \subset \left\{ (\varepsilon_1, \dots, \varepsilon_t) \in \{\pm 1\}^t : \prod_i \varepsilon_i = 1 \right\}, \quad K(\sqrt{q_1^*}, \dots, \sqrt{q_t^*}) \subset L^{\Omega(\text{Ker } \varphi)}.$$

В силу теории Галуа $\text{Gal}(L^{\Omega(\text{Ker } \varphi)}/K) \cong \text{Gal}(L/K)/\Omega(\text{Ker } \varphi) \cong \mathcal{H}_D/\text{Ker } \varphi \cong \text{Im } \varphi$, в частности $[L^{\Omega(\text{Ker } \varphi)} : K] = |\text{Im } \varphi| \leq 2^{t-1}$. В п. 3 доказано, что $[K(\sqrt{q_1^*}, \dots, \sqrt{q_t^*}) : \mathbb{Q}] = 2^t$. Отсюда $[K(\sqrt{q_1^*}, \dots, \sqrt{q_t^*}) : K] = 2^{t-1}$. Поскольку $[K(\sqrt{q_1^*}, \dots, \sqrt{q_t^*}) : K] \leq [L^{\Omega(\text{Ker } \varphi)} : K] = |\text{Im } \varphi| \leq 2^{t-1}$, это возможно только когда $|\text{Im } \varphi| = 2^{t-1}$ и $K(\sqrt{q_1^*}, \dots, \sqrt{q_t^*}) = L^{\Omega(\text{Ker } \varphi)}$. ■

Будем вычислять вместо многочлена $H_D[\theta, \alpha_*]$ его делитель

$$\hat{H}_D[\theta, \alpha_*(x) = \prod_{i: \varphi(\mathfrak{h}(A_i, B_i, C_i)) = (1, \dots, 1)} (x - \theta(\alpha_i)), \quad (16)$$

где набор $\{(A_i, B_i, C_i)\}$ является N -системой, соответствующей функции θ (т. е. удовлетворяющей условию той из теорем 2–4, которая применима к θ), а α_i — корень формы (A_i, B_i, C_i) .

Главная возникающая здесь трудность состоит в том, что $\hat{H}_D[\theta, \alpha_*]$ не инвариантен относительно $\text{Gal}(L/K)$. Из формулы (10) и того, что φ — гомоморфизм, легко видеть, что автоморфизмы из $\Omega(\text{Ker } \varphi)$ оставляют на месте $\hat{H}_D[\theta, \alpha_*(x)$; таким образом, его коэффициенты лежат в K_G . По теоремам 2–4 все числа $\theta(\alpha)$ целые алгебраические, так что коэффициенты $\hat{H}_D[\theta, \alpha_*]$ тоже целые алгебраические. Следовательно, для использования многочлена $\hat{H}_D[\theta, \alpha_*]$ в методе комплексного умножения нужно уметь восстанавливать целое алгебраическое число из K_G по его достаточно точному комплексному приближению. После получения представления коэффициентов $\hat{H}_D[\theta, \alpha_*]$ в виде точного разложения по базису \mathcal{O}_{K_G} дальнейшие действия для генерации эллиптической кривой точно такие же, как и в исходном варианте.

Следует отметить, что поле родов для генерации эллиптических кривых уже рассматривалось в 1993 г. в [4], где трудность восстановления точного значения z одного коэффициента многочлена $\hat{H}_D[\theta, \alpha_*]$ по приближённому значению преодолевается следующим образом: приближённо вычисляются все сопряжённые значения к z , после чего z ищется в виде разложения по некоторой системе образующих \mathcal{O}_{K_G} с неопределёнными целыми коэффициентами b_i ; поскольку целые коэффициенты инвариантны относительно $\text{Gal}(K_G/K)$, то все числа, сопряжённые к z , тоже выражаются через b_i , откуда можно приближённо найти систему линейных уравнений на b_i . Отметим, что этот подход требует вычисления $\theta(\alpha_i)$ для корней всех форм N -системы и всех сопряжённых многочленов к $\hat{H}_D[\theta, \alpha_*]$. Тем самым получается экономия не в количестве коэффициентов, а только в их величине.

В предлагаемом подходе требуется вычисление только многочлена $\hat{H}_D[\theta, \alpha_*]$; в частности, достаточно вычислить $\theta(\alpha_i)$ только для таких корней форм \mathfrak{a} , для которых $\varphi(\mathfrak{h}(\mathfrak{a})) = (1, \dots, 1)$. Из теоремы 13 очевидным образом следует, что таких корней в 2^{t-1} раз меньше, чем размер N -системы.

5. Оценка коэффициентов многочлена $\hat{H}_D[\theta, \alpha_*]$

В теоремах 9–12 доказано, что каждый коэффициент многочлена $\hat{H}_D[\theta, \alpha_*]$ представляется в виде $\frac{1}{2} \left(\sum_{\mu} b_{\mu} \beta_{\mu} + \sum_{\mu} b'_{\mu} \beta_{\mu}^* \right)$, где $b_{\mu}, b'_{\mu} \in \mathbb{Z}$, $\beta_{\mu} \in \mathbb{R}$, $\beta_{\mu}^* \in i\mathbb{R}$. Оценим

величины b_μ и b'_μ . Для этого сначала получим оценку на максимум модуля всех сопряжённых, т. е.

$$\left| \frac{1}{2} \tau'_\lambda \left(\sum_\mu b_\mu \beta_\mu + \sum_\mu b'_\mu \beta_\mu^* \right) \right| \leq T_0.$$

Отметим, что многочлен $\hat{H}_D[j, \alpha_*]$ не зависит от набора α_* , так что правомерно короткое обозначение $\hat{H}_D[j] = \hat{H}_D[j, \alpha_*]$.

Для получения теоретических оценок применим метод из [21].

Наряду с многочленом $\hat{H}_D[j]$ будем рассматривать также при $\varphi_0 \in \{0, 1\}^t$ многочлены

$$\hat{H}_{D, \varphi_0}[j](x) = \prod_{i: \varphi(\mathfrak{b}(A_i, B_i, C_i)) = \varphi_0} (x - j(\alpha_i)), \quad (17)$$

где, как и раньше, (A_i, B_i, C_i) пробегает представителей всех классов форм, а α_i — корень (A_i, B_i, C_i) .

По определению $\hat{H}_D[j] = \hat{H}_{D, (1, \dots, 1)}[j]$. По тем же причинам, что и для $\hat{H}_D[j]$, при всех φ_0 многочлен $\hat{H}_{D, \varphi_0}[j]$ лежит в $\mathcal{O}_{K_G}[x]$. Более того, если $\sigma \in \text{Gal}(L/\mathbb{Q})$ — автоморфизм, соответствующий классу идеалов $\mathfrak{b} \in \mathcal{H}_D$, то в силу следствия из утверждения 1 $\hat{H}_{D, \varphi_0}[j]^\sigma = \hat{H}_{D, \varphi_0 \varphi(\mathfrak{b})^{-1}}[j]$. Поскольку любой автоморфизм поля K_G продолжается до элемента из $\text{Gal}(L/\mathbb{Q})$, то для любого $\tau \in \text{Gal}(K_G/\mathbb{Q})$ найдётся $\varphi_0 = \varphi_0(\tau)$ (фиксируем одно из продолжений), такое, что $\hat{H}_{D, \varphi_1}[j]^\tau = \hat{H}_{D, \varphi_1 \varphi_0(\tau)}[j]$ для любого $\varphi_1 \in \{\pm 1\}^t$.

Теорема 14. Каждый коэффициент многочлена $\hat{H}_{D, \varphi_0}[j]$ по модулю не превосходит

$$\begin{aligned} & \exp \left(c_5 h + c_1 N \left(\ln^2 N + 4\gamma \ln N + c_6 + \frac{\ln N + \gamma + 1}{N} \right) \right) \leq \\ & \leq \exp(c_1 N \ln^2 N + c_2 N \ln N + c_3 N + c_1 \ln N + c_4) = T_0, \end{aligned}$$

где $N = \sqrt{|D|/3}$; $\gamma = 0,577\dots$ — константа Эйлера; $c_1 = \sqrt{3}\pi = 5,441\dots$; $c_2 = 18,587\dots$; $c_3 = 17,442\dots$; $c_4 = 11,594\dots$; $c_5 = 3,011\dots$; $c_6 = 2,566\dots$ Асимптотическая верхняя оценка

$$T_0 = \exp O \left(\sqrt{|D|} \ln^2 |D| \right)$$

также выполняется для других функций θ .

Доказательство. Будем следовать [21, Section 4].

В произведении из (17) можно считать, что (A_i, B_i, C_i) — приведённые формы (поскольку замена формы на эквивалентную соответствует некоторому $SL_2(\mathbb{Z})$ -преобразованию корня формы, а j инвариантен относительно таких преобразований). Оценим для приведённой формы (A, B, C) значение $j \left(\frac{-B + \sqrt{D}}{2A} \right)$. Аргумент функции j находится в области $\{z \in \mathbb{H} : |z| \geq 1, |\text{Re } z| \leq 1/2\}$. Следовательно, $\text{Im } z \geq \sqrt{3}/2$ и $|q| = |e^{2\pi iz}| \leq e^{-\pi\sqrt{3}}$. Далее,

$$j(z) = \frac{1}{q} + 744 + \sum_{m=1}^{\infty} c_m q^m,$$

где коэффициенты ряда Фурье, согласно [22], оцениваются как $|c_m| \leq \frac{e^{4\pi\sqrt{m}}}{\sqrt{2m^{3/4}}}$. Следовательно,

$$\left| j \left(\frac{-B + \sqrt{D}}{2A} \right) - \frac{1}{q} \right| \leq 744 + \sum_{m=1}^{\infty} \frac{e^{4\pi\sqrt{m}}}{\sqrt{2m^{3/4}}} e^{-\pi\sqrt{3}m} = k_1 = 2114,566\dots$$

и $\left| j \left((-B + \sqrt{D}) / (2A) \right) \right| \leq 1/|q| + k_1 \leq k_2/|q|$, где $k_2 = 1 + k_1 e^{-\pi\sqrt{3}} = 10,163\dots$

Перенумеруем приведённые формы так, чтобы $\{(A_i, B_i, C_i) : 1 \leq i \leq \deg \hat{H}_D[j]\}$ — все приведённые формы, по которым вычисляется произведение (17), — были расположены в порядке возрастания $|1/q_i| = e^{\pi\sqrt{|D|}/A_i}$. Тогда коэффициент при x^k многочлена $\hat{H}_{D,\varphi_0}[j]$ по модулю не превосходит

$$C_{\deg \hat{H}_D[j]}^k \prod_{i=k+1}^{\deg \hat{H}_D[j]} \frac{k_2}{|q_i|} \leq (2k_2)^{h/2^{t-1}} \prod_{i=1}^{h/2^{t-1}} e^{\pi\sqrt{|D|}/A_i}.$$

Следовательно, логарифм любого коэффициента $\hat{H}_{D,\varphi_0}[j]$ не превосходит

$$\frac{h}{2^{t-1}} \ln(2k_2) + \pi\sqrt{|D|} \sum_{i=1}^{h/2^{t-1}} \frac{1}{A_i} \leq h \ln(2k_2) + \pi\sqrt{|D|} \sum_{i=1}^h \frac{1}{A_i}.$$

Оценка для последней суммы, вычисленная в [21, Theorem 1.2], завершает доказательство для j . Оценки для других функций θ следуют из доказанной оценки и [23, Proposition 3]. ■

На практике будем использовать эвристические, но более точные оценки.

В [23] в качестве эвристической оценки сверху, достаточно близкой к точному значению, для логарифмов модулей коэффициентов многочлена $H_D[j]$ предлагается сумма

$$\pi\sqrt{|D|} \sum_{(A,B,C)} \frac{1}{A},$$

где сумма берётся по всем приведённым формам. Там же в качестве аналогичной оценки, соответствующей θ вместо j , предлагается эта же сумма, умноженная на некоторую константу, зависящую только от выбранного инварианта θ , а именно на отношение степеней $\deg_j \Phi / \deg_\theta \Phi$, где многочлен от двух переменных Φ связывает функции θ и j так, что $\Phi(\theta(z), j(z)) = 0$.

Тривиальные изменения рассуждений из [23] применительно к многочлену $\hat{H}_D[j]$ дают эвристическую оценку

$$\ln T_0 \sim \pi\sqrt{|D|} \max_{\varepsilon \in \{\pm 1\}^t} \sum_{(A,B,C): \varphi(\mathfrak{h}(A,B,C)) = \varepsilon} \frac{1}{A} \quad (18)$$

при использовании инварианта j . При использовании других функций θ следует умножить оценку на $\deg_j \Phi / \deg_\theta \Phi$.

Пусть $z = \left(\sum_{\mu} b_{\mu} \beta_{\mu} + \sum_{\mu} b'_{\mu} \beta_{\mu}^* \right) / 2$ — коэффициент многочлена $\hat{H}_{D,\varphi_0}[j]$, $b_{\mu}, b'_{\mu} \in \mathbb{Z}$.

Как было отмечено выше, действие группы $\text{Gal}(K_G/\mathbb{Q})$ переводит многочлен $\hat{H}_{D,\varphi_1}[j]$ в многочлен того же вида, поэтому для любого $\lambda \in \{0, 1\}^t$ справедливо неравенство

$$|\tau'_{\lambda}(z)| \leq T_0.$$

6. Построение рациональных приближений к базису кольца целых

Для построения совместных приближений существуют различные универсальные алгоритмы, изучению которых посвящена работа [24]; на практике характеристики получаемых приближений существенно различаются для разных алгоритмов, для практических целей, по-видимому, наилучшим является алгоритм скалярных произведений из [24, Chapter 6A]. К сожалению, для универсальных алгоритмов довольно трудно получить теоретические оценки качества. Поэтому мы предлагаем алгоритм, подходящий только для наборов чисел нужного вида, работающий на практике примерно столь же хорошо, сколь и алгоритм скалярных произведений, и допускающий теоретические оценки.

Основная часть следующей теоремы по существу содержится в работе [25]. Основные её отличия от рассуждений в [25] заключаются в явной формулировке (в том числе явных константах), введении функции \mathfrak{M} ([25] оперирует двойственными базами, что эквивалентно $\mathfrak{M} = 1$) и специализацией для интересующего нас случая ([25] не требует нормальности расширения M/\mathbb{Q} и содержит также обратную теорему).

Теорема 15. Пусть $M \subset \mathbb{R}$ — поле, такое, что M/\mathbb{Q} — расширение Галуа степени m . Пусть W_1, \dots, W_m и W_1^*, \dots, W_m^* — два \mathbb{Q} -базиса M и $\mathfrak{M} : \text{Gal}(M/\mathbb{Q}) \rightarrow \mathbb{R}$ — функция (необязательно гомоморфизм), такие, что для всех $1 \leq l, l' \leq m$ выполнено равенство

$$\sum_{\tau \in \text{Gal}(M/\mathbb{Q})} \mathfrak{M}(\tau) \tau(W_l W_{l'}^*) = \begin{cases} 1, & \text{если } l = l', \\ 0, & \text{если } l \neq l'. \end{cases}$$

Обозначим

$$C = \sum_{\substack{\tau \in \text{Gal}(M/\mathbb{Q}) \\ \tau \neq Id}} |\mathfrak{M}(\tau) \tau(W_1)|$$

и при $i = 2, \dots, m$

$$C_i = \sum_{\substack{\tau \in \text{Gal}(M/\mathbb{Q}) \\ \tau \neq Id}} \left| \mathfrak{M}(\tau) \left(\tau(W_i) - W_i \frac{\tau(W_1)}{W_1} \right) \right|.$$

Пусть также положительное число Δ и целые числа $\Lambda_1, \dots, \Lambda_m$ таковы, что

$$\sum_{i=1}^m \Lambda_i W_i^* = Z \geq 1, \quad \left| \tau \left(\sum_{i=1}^m \Lambda_i W_i^* \right) \right| \leq \frac{\Delta}{Z^{\frac{1}{m-1}}} \quad \text{для всех } \tau \in \text{Gal}(M/\mathbb{Q}), \tau \neq Id.$$

Тогда

- справедливо неравенство $|\Lambda_1| \geq |\mathfrak{M}(Id) W_1| Z - C \Delta$;
- если $|\Lambda_1| > C \Delta$, то $\mathfrak{M}(Id) \neq 0$ и при всех $i = 2, \dots, m$ справедлива оценка

$$\left| \frac{\Lambda_i}{\Lambda_1} - \frac{W_i}{W_1} \right| \leq C_i \frac{\Delta}{|\Lambda_1| \left(\frac{|\Lambda_1| - C \Delta}{|\mathfrak{M}(Id) W_1|} \right)^{\frac{1}{m-1}}}.$$

Доказательство. По условию для каждого $l = 1, \dots, m$ справедливо равенство

$$\begin{aligned} \Lambda_l &= \sum_{l'=1}^m \Lambda_{l'} \left(\sum_{\tau \in \text{Gal}(M/\mathbb{Q})} \mathfrak{M}(\tau) \tau(W_l W_{l'}^*) \right) = \sum_{\tau \in \text{Gal}(M/\mathbb{Q})} \mathfrak{M}(\tau) \tau(W_l) \left(\sum_{l'=1}^m \Lambda_{l'} \tau(W_{l'}^*) \right) = \\ &= \mathfrak{M}(Id) W_l Z + \sum_{\substack{\tau \in \text{Gal}(M/\mathbb{Q}) \\ \tau \neq Id}} \mathfrak{M}(\tau) \tau(W_l) \tau(Z). \end{aligned} \quad (19)$$

Подставим $l = 1$:

$$\Lambda_1 = \mathfrak{M}(Id)W_1Z + \sum_{\substack{\tau \in \text{Gal}(M/\mathbb{Q}) \\ \tau \neq Id}} \mathfrak{M}(\tau)\tau(W_1)\tau(Z). \quad (20)$$

Переносим первое слагаемое из правой части в левую и используя определение C и оценку $\tau(Z)$, находим

$$|\Lambda_1 - \mathfrak{M}(Id)W_1Z| \leq \frac{C\Delta}{Z^{\frac{1}{m-1}}} \leq C\Delta.$$

Это доказывает первое утверждение теоремы.

Пусть $|\Lambda_1| > C\Delta$. Тогда $|\mathfrak{M}(Id)W_1Z| \geq |\Lambda_1| - C\Delta$. Следовательно, $\mathfrak{M}(Id) \neq 0$ и

$$Z \geq \frac{|\Lambda_1| - C\Delta}{|\mathfrak{M}(Id)W_1|}. \quad (21)$$

Вычтем из равенства (19) равенство (20), умноженное на W_l/W_1 , и воспользуемся определением C_l и оценкой $\tau(Z)$:

$$\left| \Lambda_l - \frac{W_l}{W_1} \Lambda_1 \right| \leq C_l \frac{\Delta}{Z^{\frac{1}{m-1}}}.$$

Разделим полученное неравенство на $|\Lambda_1|$:

$$\left| \frac{\Lambda_l}{\Lambda_1} - \frac{W_l}{W_1} \right| \leq C_l \frac{\Delta}{|\Lambda_1| Z^{\frac{1}{m-1}}}.$$

Применение оценки (21) завершает доказательство. ■

В работе [25] используется знание структуры группы единиц \mathcal{O}_M^* (которое даёт теорема Дирихле) и находится число $\sum_{i=1}^m \Lambda_i W_i^*$ как единица специального вида. Это позволяет доказать интересные теоретические результаты, но довольно неудобно с практической точки зрения. Используем другой подход.

Будем строить совместные приближения к элементам поля $\mathcal{M} = K_G \cap \mathbb{R}$, для этого применим теорему 15 к полю $M = \mathcal{M}$. Соответственно $m = [\mathcal{M} : \mathbb{Q}] = 2^{t-1}$, $t \geq 2$, и $\text{Gal}(\mathcal{M}/\mathbb{Q})$ состоит из автоморфизмов τ_λ при $\lambda \in \{0, 1\}^{t-1}$, определённых в (15).

Наборы, связанные с полем \mathcal{M} , удобно нумеровать векторами из множества $\{0, 1\}^{t-1}$. Поэтому дальше будем полагать, что заданы два \mathbb{Q} -базиса поля \mathcal{M} , ω_μ и ω_μ^* при $\mu \in \{0, 1\}^{t-1}$, а также функция $\mathfrak{M} : \text{Gal}(\mathcal{M}/\mathbb{Q}) \rightarrow \mathbb{R}$, удовлетворяющие следующим условиям:

- 1) $\omega_{0, \dots, 0}^* = 1$;
- 2) любой элемент кольца целых \mathcal{O}_M поля \mathcal{M} разлагается по базису $\{\omega_\mu^*\}$ с целыми коэффициентами;
- 3) если $\lambda, \lambda' \in \{0, 1\}^{t-1}$, то

$$\sum_{\mu \in \{0, 1\}^{t-1}} \mathfrak{M}(\tau_\mu) \tau_\mu(\omega_\lambda \omega_{\lambda'}^*) = \begin{cases} 1, & \text{если } \lambda = \lambda', \\ 0, & \text{если } \lambda \neq \lambda'. \end{cases} \quad (22)$$

Будем называть такую пару \mathfrak{M} -парой. Легко видеть, что последнее условие обеспечивает выполнение условия на базисы теоремы 15, применённой к числам

$$\begin{aligned} W_{1+\mu_1+2\mu_2+2^2\mu_3+\dots+2^{t-2}\mu_{t-1}} &= \omega_\mu, \\ W_{1+\mu_1+2\mu_2+2^2\mu_3+\dots+2^{t-2}\mu_{t-1}}^* &= \omega_\mu^*. \end{aligned}$$

Заметим, что если $x \in \mathcal{O}_M$, то $x\beta_{0,\dots,0}^* \in \mathcal{O}_{K_G} \cap i\mathbb{R}$. Из теорем 9–12 легко видеть, что справедливы следующие следствия. Как и в самих теоремах, для \sqrt{d} выбирается то из двух значений, которое соответствует произведению $\sqrt{q_1^*} \cdot \dots \cdot \sqrt{q_t^*}$.

Следствие 1. Если

$$\begin{aligned} \omega_{\mu_1,\dots,\mu_{t-1}} &= \frac{\beta_{\mu_1,\dots,\mu_{t-1}}}{\beta_{0,\dots,0}}, \\ \omega_{\mu_1,\dots,\mu_{t-1}}^* &= \frac{\beta_{\mu_1,\dots,\mu_{t-1}}^*}{\beta_{0,\dots,0}^*}, \\ \mathfrak{M}(\tau_{\mu_1,\dots,\mu_{t-1}}) &= (-1)^{\mu_1+\dots+\mu_{t-1}} \frac{\tau_{\mu_1,\dots,\mu_{t-1}} (\beta_{0,\dots,0}\beta_{0,\dots,0}^*)}{\sqrt{d}}, \end{aligned} \quad (23)$$

то условия 1–3 выполнены.

Следствие 2. Если

$$\begin{aligned} \omega_{\mu_1,\dots,\mu_{t-1}} &= \frac{\beta_{\mu_1,\dots,\mu_{t-1}}^*}{\beta_{0,\dots,0}^*}, \\ \omega_{\mu_1,\dots,\mu_{t-1}}^* &= \frac{\beta_{\mu_1,\dots,\mu_{t-1}}}{\beta_{0,\dots,0}}, \\ \mathfrak{M}(\tau_{\mu_1,\dots,\mu_{t-1}}) &= (-1)^{\mu_1+\dots+\mu_{t-1}} \frac{\tau_{\mu_1,\dots,\mu_{t-1}} (\beta_{0,\dots,0}\beta_{0,\dots,0}^*)}{\sqrt{d}}, \end{aligned} \quad (24)$$

то условия 1–3 выполнены.

Помимо базисов W_i, W_i^* и функции \mathfrak{M} , в теореме 15 фигурируют набор целых чисел Λ_i и константа Δ . Далее опишем алгоритм построения набора A_μ , такого, что числа $\Lambda_{1+\mu_1+2\mu_2+2^2\mu_3+\dots+2^{t-2}\mu_{t-1}} = A_{\mu_1,\dots,\mu_{t-1}}$ удовлетворяют условию теоремы 15 с некоторым Δ .

Пусть $\lambda \in \{0, 1\}^{t-1}$, $\lambda \neq (0, \dots, 0)$. Положим $\delta_\lambda = (q_1^*)^{\lambda_1} \dots (q_{t-1}^*)^{\lambda_{t-1}} (q_t^*)^{\lambda_{u+1} \oplus \dots \oplus \lambda_{t-1}}$. Если δ_λ чётно, положим $g_\lambda = \sqrt{\delta_\lambda}/2$, в противном случае $-g_\lambda = (1 + \sqrt{\delta_\lambda})/2$. Тогда $g_\lambda \in \mathcal{O}_M$.

Будем использовать цепные дроби. Напомним, что для любого числа $X \in \mathbb{R}$ определена последовательность его полных частных X_0, X_1, X_2, \dots и неполных частных a_0, a_1, a_2, \dots следующим образом: $X_0 = X$, $a_n = \lfloor X_n \rfloor$, $X_{n+1} = 1/(X_n - a_n)$, причём последовательность конечна (т.е. X_n не определено при некотором n) тогда и только тогда, когда $X \in \mathbb{Q}$. Кроме того, определена последовательность подходящих дробей $\frac{P_0}{Q_0}, \frac{P_1}{Q_1}, \frac{P_2}{Q_2}, \dots$ к X вместе с формальным выражением $\frac{P_{-1}}{Q_{-1}}$ следующим образом: $P_{-1} = 1, Q_{-1} = 0, P_0 = a_0, Q_0 = 1, P_{n+1} = a_{n+1}P_n + P_{n-1}, Q_{n+1} = a_{n+1}Q_n + Q_{n-1}$. Хорошо известно (например, [26, Теорема 9 и Теорема 12]), что при $n \geq 0$

$$\left| X - \frac{P_n}{Q_n} \right| < \frac{1}{Q_n Q_{n+1}}, \text{ если } X_{n+2} \text{ определено;} \quad (25)$$

$$Q_n \geq 2^{\frac{n-1}{2}}. \quad (26)$$

Будем рассматривать случай квадратичных иррациональностей; воспользуемся некоторыми результатами из [27, §II.10], собранными в следующем утверждении.

Утверждение 2. Пусть a, b, c — целые числа, взаимно простые в совокупности, и $\delta = b^2 - ac > 0$ не равно полному квадрату. Корни уравнения $ax^2 + 2bx + c = 0$ будем называть иррациональностями определителя δ .

Пусть $X = (-b + \sqrt{\delta})/a$ — иррациональность определителя δ . Тогда все полные частные X тоже являются иррациональностями определителя δ и имеют вид $X_n = (x_n + \sqrt{\delta})/y_n$, где $x_n, y_n \in \mathbb{Z}$ определены единственным образом. Если $a_n = [X_n]$ — неполные частные для X и $y_{-1} = -c = (\delta - b^2)/a \in \mathbb{Z}$, то справедливы следующие формулы:

$$\begin{aligned} x_n &= y_{n-1}a_{n-1} - x_{n-1}, & n \geq 1, \\ \delta &= x_n^2 + y_n y_{n-1}, & n \geq 0, \\ y_n &= y_{n-2} - a_{n-1}(x_n - x_{n-1}), & n \geq 1. \end{aligned} \quad (27)$$

Кроме того, при $n \geq 0$

$$X_1 \dots X_n = \frac{(-1)^n}{P_{n-1} - Q_{n-1}X}; \quad aP_{n-1}^2 + 2bP_{n-1}Q_{n-1} + cQ_{n-1}^2 = (-1)^n y_n.$$

Назовём число $(x + \sqrt{\delta})/y$, $x, y \in \mathbb{Z}$, приведённым, если $(x + \sqrt{\delta})/y > 1$ и $-1 < (x - \sqrt{\delta})/y < 0$. Число $(x + \sqrt{\delta})/y$ приведённое тогда и только тогда, когда $0 < \sqrt{\delta} - x < y < \sqrt{\delta} + x$. Если X приведённое, то все его полные частные X_n тоже приведённые.

Введём несколько необходимых определений.

Будем параллельно строить цепные дроби для всех чисел g_λ , $\lambda \in \{0, 1\}^{t-1}$, $\lambda \neq 0$. Полные частные для g_λ будем обозначать $X_{\lambda,n}$, неполные частные — $a_{\lambda,n}$, числители и знаменатели подходящих дробей — $P_{\lambda,n}$ и $Q_{\lambda,n}$. Величины x_n, y_n из утверждения 2, построенные для $X = g_\lambda$, будем обозначать $x_{\lambda,n}$ и $y_{\lambda,n}$. Пусть σ_λ обозначает единственный нетривиальный автоморфизм поля $\mathbb{Q}(g_\lambda)$.

Если δ_λ нечётно, то по определению g_λ есть иррациональность определителя δ_λ , $x_{\lambda,0} = 1$, $y_{\lambda,0} = 2$, $y_{\lambda,-1} = (\delta_\lambda - 1)/2$. По индукции из (27) легко видеть, что $x_{\lambda,n}$ нечётно при всех n и $y_{\lambda,n}$ чётно при всех n ; обозначим $x'_{\lambda,n} = (x_{\lambda,n} - 1)/2 \in \mathbb{Z}$ и $y'_{\lambda,n} = y_{\lambda,n}/2 \in \mathbb{Z}$. Квадратный трёхчлен $ax^2 + 2bx + c$, где a, b, c определены утверждением 2, имеет старший коэффициент 2 и корни $g_\lambda, \sigma_\lambda(g_\lambda)$. Следовательно, последнее равенство утверждения 2 можно переписать как $2(P_{\lambda,n-1} - Q_{\lambda,n-1}g_\lambda)\sigma_\lambda(P_{\lambda,n-1} - Q_{\lambda,n-1}g_\lambda) = (-1)^n y_{\lambda,n} = (-1)^n 2y'_{\lambda,n}$.

Если δ_λ чётно, то по определению g_λ есть иррациональность определителя $\delta_\lambda/4$, $x_{\lambda,0} = 0$, $y_{\lambda,0} = 1$, $y_{\lambda,-1} = \delta_\lambda/4$. Обозначим $x'_{\lambda,n} = x_{\lambda,n}$ и $y'_{\lambda,n} = y_{\lambda,n}$. Квадратный трёхчлен $ax^2 + 2bx + c$, где a, b, c определены утверждением 2, имеет старший коэффициент 1 и корни $g_\lambda, \sigma_\lambda(g_\lambda)$. Следовательно, последнее равенство утверждения 2 можно переписать как $(P_{\lambda,n-1} - Q_{\lambda,n-1}g_\lambda)\sigma_\lambda(P_{\lambda,n-1} - Q_{\lambda,n-1}g_\lambda) = (-1)^n y_{\lambda,n} = (-1)^n y'_{\lambda,n}$.

В обоих случаях

$$X_{\lambda,n} = \frac{g_\lambda + x'_{\lambda,n}}{y'_{\lambda,n}},$$

$$(P_{\lambda,n-1} - Q_{\lambda,n-1}g_\lambda)\sigma_\lambda(P_{\lambda,n-1} - Q_{\lambda,n-1}g_\lambda) = (-1)^n y'_{\lambda,n}. \quad (28)$$

Утверждение 2 даёт эффективный способ последовательного построения чисел $x'_{\lambda,n}$, $y'_{\lambda,n}$, $a_{\lambda,n} = [X_{\lambda,n}]$, а по ним $P_{\lambda,n}$ и $Q_{\lambda,n}$. Для алгоритма понадобятся числа $x'_{\lambda,n}$, $y'_{\lambda,n}$ и

$$z_{\lambda,n} = (X_{\lambda,1} \dots X_{\lambda,n})^{-1} = (-1)^n (P_{\lambda,n-1} - Q_{\lambda,n-1}g_\lambda) \in \mathcal{O}_M. \quad (29)$$

Отсюда и из (28) следует, что для любого $n \geq 0$

$$z_{\lambda,n} \sigma_{\lambda}(z_{\lambda,n}) = (-1)^n y'_{\lambda,n}. \quad (30)$$

Числа A_{μ} будем получать, пересчитывая разложение произведения целых алгебраических чисел $\prod_{\lambda \neq 0} ((-1)^{n_{\lambda}} \sigma_{\lambda}(z_{\lambda,n_{\lambda}})) = \sum_{\mu} A_{\mu} \omega_{\mu}^*$ по базису ω_{μ}^* . В силу условия 2 \mathfrak{M} -пары и (29) числа A_{μ} будут целыми.

На каждом шаге алгоритм увеличивает ровно одно из чисел n_{λ} . На $\sum_{\mu} A_{\mu} \omega_{\mu}^*$ этот переход действует умножением на

$$\frac{(-1)^{n+1} \sigma_{\lambda}(z_{\lambda,n+1})}{(-1)^n \sigma_{\lambda}(z_{\lambda,n})} = \frac{y'_{\lambda,n+1}/z_{\lambda,n+1}}{y'_{\lambda,n}/z_{\lambda,n}} = \frac{X_{\lambda,n+1} y'_{\lambda,n+1}}{y'_{\lambda,n}} = \frac{g_{\lambda} + x'_{\lambda,n+1}}{y'_{\lambda,n}}.$$

Таким образом, нужно переходить от набора чисел A_{μ} к набору чисел A'_{μ} , такому, что

$$\sum_{\mu} A'_{\mu} \omega_{\mu}^* = \left(\sum_{\xi} A_{\xi} \omega_{\xi}^* \right) \frac{g_{\lambda} + x_{\lambda}}{y_{\lambda}},$$

где $x_{\lambda} = x'_{\lambda,n_{\lambda+1}}$ и $y_{\lambda} = y'_{\lambda,n_{\lambda}}$. Поскольку $\{\omega_{\mu}^*\}$ образуют \mathbb{Q} -базис поля \mathcal{M} и числа g_{μ} лежат в этом поле, то в самом начале можно предвычислить такие $c_{\mu\xi\eta} \in \mathbb{Q}$, что

$$\omega_{\xi}^* g_{\eta} = \sum_{\mu} c_{\mu\xi\eta} \omega_{\mu}^*.$$

Зная числа $c_{\mu\xi\eta}$, можем вычислить

$$\left(\sum_{\xi} A_{\xi} \omega_{\xi}^* \right) \frac{g_{\lambda} + x_{\lambda}}{y_{\lambda}} = \frac{1}{y_{\lambda}} \left(\sum_{\xi} A_{\xi} \sum_{\mu} c_{\mu\xi\lambda} \omega_{\mu}^* + \sum_{\xi} A_{\xi} \omega_{\xi}^* x_{\lambda} \right) = \sum_{\mu} \frac{\sum_{\xi} A_{\xi} c_{\mu\xi\lambda} + A_{\mu} x_{\lambda}}{y_{\lambda}} \omega_{\mu}^*.$$

Теперь можно предъявить алгоритм.

Алгоритм построения совместных приближений. Входные данные — набор δ_{λ} , g_{λ} , $c_{\mu\xi\eta}$, введённых выше, а также порог $N_0 > 0$. Выходные данные — целые числа A_{μ} , такие, что $|A_{0,\dots,0}| \geq N_0$ и $A_{\mu}/A_{0,\dots,0}$ — приближение к $\omega_{\mu}/\omega_{0,\dots,0}$ для каждого $\mu \in \{0, 1\}^{t-1}$.

Алгоритм в процессе работы хранит набор из 2^{t-1} целых чисел A_{μ} , а также вспомогательные наборы целых неотрицательных чисел x_{λ} , натуральных чисел $(y_{\lambda}, \tilde{y}_{\lambda})$ и вещественных положительных чисел $(z_{\lambda}, \tilde{z}_{\lambda})$, где $\lambda \in \{0, 1\}^{t-1}$, $\lambda \neq (0, \dots, 0)$. Эти наборы имеют следующий смысл: если после некоторого числа итераций на шаге 3 (см. ниже) значение λ было выбрано n_{λ} раз, то в определённых выше обозначениях $x_{\lambda} = x'_{\lambda,n_{\lambda}}$, $(y_{\lambda}, \tilde{y}_{\lambda}) = (y'_{\lambda,n_{\lambda}}, y'_{\lambda,n_{\lambda}-1})$, $(z_{\lambda}, \tilde{z}_{\lambda}) = (z_{\lambda,n_{\lambda}}, z_{\lambda,n_{\lambda}-1})$, $\sum_{\mu} A_{\mu} \omega_{\mu}^* = \prod_{\lambda \neq 0} ((-1)^{n_{\lambda}} \sigma_{\lambda}(z_{\lambda,n_{\lambda}}))$.

Действия алгоритма.

- 1) *Инициализация.* Присвоить начальные значения: $A_{0,\dots,0} := 1$; $A_{\lambda} := 0$; $x_{\lambda} := 0$; $(y_{\lambda}, \tilde{y}_{\lambda}) := (1, \lfloor \delta_{\lambda}/4 \rfloor)$; $(z_{\lambda}, \tilde{z}_{\lambda}) := (1, g_{\lambda})$ для всех $\lambda \in \{0, 1\}^{t-1}$, $\lambda \neq (0, \dots, 0)$.
- 2) *Итерации.* Пока $|A_{0,\dots,0}| < N_0$, повторять следующие шаги.
- 3) Выбрать некоторое λ , такое, что $z_{\lambda} = \max_{\mu \neq (0,\dots,0)} z_{\mu}$.
- 4) Вычислить $a = \lfloor (g_{\lambda} + x_{\lambda})/y_{\lambda} \rfloor$.
- 5) Присвоить $(z_{\lambda}, \tilde{z}_{\lambda}) := (\tilde{z}_{\lambda} - a z_{\lambda}, z_{\lambda})$.

6) Запомнить $x = x_\lambda$, присвоить $x_\lambda := ay_\lambda - x_\lambda - 4\{\delta_\lambda/4\}$, после чего присвоить $(y_\lambda, \tilde{y}_\lambda) := (\tilde{y}_\lambda - a(x_\lambda - x), y_\lambda)$. (Как будет показано, новое значение x_λ всегда целое неотрицательное, а новое значение y_λ натуральное.)

7) Вычислить $A'_\mu = \left(\sum_\xi A_\xi c_{\mu\xi} + A_\mu x_\lambda \right) / \tilde{y}_\lambda$ для всех μ . (Как было показано, $A'_\mu \in \mathbb{Z}$ для всех μ .) Присвоить $A_\mu := A'_\mu$.

Доказательство следующей теоремы есть в [28].

Теорема 16. Алгоритм завершается за $O(\ln N_0)$ шагов. В процессе его работы всегда выполнены неравенства $0 \leq x_\lambda < \sqrt{\delta_\lambda} - g_\lambda$; $0 < y_\lambda < \sqrt{\delta_\lambda}$; $Z = \sum_\mu A_\mu \omega_\mu^* \geq 1$;

$$\left| \tau_\lambda \left(\sum_\mu A_\mu \omega_\mu^* \right) \right| \leq \frac{\sqrt{|d|}^m}{Z^{m-1}} \text{ при } \lambda \neq (0, \dots, 0).$$

Доказательство. Начнём с оценок величин $x'_{\lambda,n}$, $y'_{\lambda,n}$.

Лемма 5. Пусть $\lambda \in \{0, 1\}^{t-1}$, $\lambda \neq 0$; $n \geq 1$ — целое число. Тогда

$$0 \leq x'_{\lambda,n} < \sqrt{\delta_\lambda} - g_\lambda, \quad 0 < y'_{\lambda,n} < \sqrt{\delta_\lambda}, \quad -1 < \sigma_\lambda(X_{\lambda,n}) < 0.$$

Доказательство. Будем отдельно рассматривать случаи чётного и нечётного δ_λ .

Пусть δ_λ нечётно. По определению $X_{\lambda,1} = 1/(g_\lambda - \lfloor g_\lambda \rfloor)$. Очевидно, что $X_{\lambda,1} > 1$. Кроме того, $\sigma_\lambda(X_{\lambda,1}) = 1/(1 - g_\lambda - \lfloor g_\lambda \rfloor)$ и, поскольку $g_\lambda > 1$, получаем $-1 < \sigma_\lambda(X_{\lambda,1}) < 0$. Следовательно, в силу утверждения 2 все полные частные для g_λ , начиная с $X_{\lambda,1}$, являются приведёнными иррациональностями определителя δ_λ , т. е. $0 < \sqrt{\delta_\lambda} - x_{\lambda,n} < y_{\lambda,n} < \sqrt{\delta_\lambda} + x_{\lambda,n}$ при $n \geq 1$. Поскольку в этом случае $x'_{\lambda,n} = (x_{\lambda,n} - 1)/2$ и $y'_{\lambda,n} = (y_{\lambda,n})/2$, получаем заявленные оценки.

Пусть δ_λ чётно. Как и в предыдущем случае, $X_{\lambda,1} = 1/(g_\lambda - \lfloor g_\lambda \rfloor) > 1$. Кроме того, $\sigma_\lambda(X_{\lambda,1}) = -1/(g_\lambda + \lfloor g_\lambda \rfloor)$ и, поскольку $g_\lambda > 1$, получаем $-1 < \sigma_\lambda(X_{\lambda,1}) < 0$. Следовательно, в силу утверждения 2 все полные частные для g_λ , начиная с $X_{\lambda,1}$, являются приведёнными иррациональностями определителя $\delta_\lambda/4$, т. е. $0 < \sqrt{\delta_\lambda}/2 - x_{\lambda,n} < y_{\lambda,n} < \sqrt{\delta_\lambda}/2 + x_{\lambda,n}$ при $n \geq 1$. Поскольку в этом случае $x'_{\lambda,n} = x_{\lambda,n}$ и $y'_{\lambda,n} = y_{\lambda,n}$, получаем заявленные оценки. ■

Поскольку $X_{\lambda,n} = (g_\lambda + x'_{\lambda,n})/y'_{\lambda,n}$, немедленно получаем

Следствие. При $n \geq 1$ справедлива оценка

$$X_{\lambda,n} < \sqrt{\delta_\lambda}. \quad (31)$$

Напомним, что n_λ при $\lambda \neq 0$ обозначает количество выборов λ на шаге 3 алгоритма.

Неравенство $Z = \prod_{\lambda \neq 0} ((-1)^{n_\lambda} \sigma_\lambda(z_{\lambda,n_\lambda})) \geq 1$ немедленно следует из последнего нера-

венства леммы 5 и определения $z_{\lambda,n_\lambda} = 1/(X_{\lambda,1} \cdot \dots \cdot X_{\lambda,n_\lambda})$.

Лемма 6.

$$\frac{\max_{\mu \neq (0, \dots, 0)} z_{\mu, n_\mu}}{\min_{\mu \neq (0, \dots, 0)} z_{\mu, n_\mu}} \leq \sqrt{|d|}.$$

Доказательство. Перед итерациями левая часть равна 1, так что неравенство верно. Допустим, что после некоторого числа итераций неравенство верно, и предположим, что на очередной итерации алгоритм выбрал значение λ на шаге 3, т. е. $z_{\lambda, n_\lambda} = \max_{\mu \neq (0, \dots, 0)} z_{\mu, n_\mu}$. Обозначим $n'_\lambda = n_\lambda + 1$ и $n'_\mu = n_\mu$ при $\mu \neq \lambda$, $\mu \neq (0, \dots, 0)$.

Очевидно, $X_{\lambda, n'_\lambda} > 1$, так что $z_{\lambda, n'_\lambda} < z_{\lambda, n_\lambda}$. Возможны два случая:

- 1) $z_{\lambda, n'_\lambda} \geq \min_{\mu \neq (0, \dots, 0)} z_{\mu, n_\mu}$. В этом случае $\min_{\mu \neq (0, \dots, 0)} z_{\mu, n'_\mu} = \min_{\mu \neq (0, \dots, 0)} z_{\mu, n_\mu}$, следовательно,

$$\frac{\max_{\mu \neq (0, \dots, 0)} z_{\mu, n'_\mu}}{\min_{\mu \neq (0, \dots, 0)} z_{\mu, n'_\mu}} \leq \frac{\max_{\mu \neq (0, \dots, 0)} z_{\mu, n_\mu}}{\min_{\mu \neq (0, \dots, 0)} z_{\mu, n_\mu}} \leq \sqrt{|d|}.$$
- 2) $z_{\lambda, n'_\lambda} < \min_{\mu \neq (0, \dots, 0)} z_{\mu, n_\mu}$. В этом случае $\min_{\mu \neq (0, \dots, 0)} z_{\mu, n'_\mu} = z_{\lambda, n'_\lambda}$; используя (31), находим

$$\frac{\max_{\mu \neq (0, \dots, 0)} z_{\mu, n'_\mu}}{\min_{\mu \neq (0, \dots, 0)} z_{\mu, n'_\mu}} \leq \frac{z_{\lambda, n_\lambda}}{z_{\lambda, n'_\lambda}} = X_{\lambda, n_\lambda+1} < \sqrt{\delta_\lambda} \leq \sqrt{|d|}.$$

Лемма доказана. ■

Напомним, что σ_λ — автоморфизм поля $\mathbb{Q}(g_\lambda) \subset \mathcal{M}$. Заметим, что при любых λ и μ отображение τ_μ можно ограничить на поле $\mathbb{Q}(g_\lambda)$. Поскольку

$$\begin{aligned} \tau_\mu \left(\sqrt{(q_1^*)^{\lambda_1} \dots (q_{t-1}^*)^{\lambda_{t-1}} (q_t^*)^{\lambda_{u+1} \oplus \dots \oplus \lambda_{t-1}}} \right) &= \\ &= ((-1)^{\mu_1} \sqrt{q_1^*})^{\lambda_1} \dots ((-1)^{\mu_{t-1}} \sqrt{q_{t-1}^*})^{\lambda_{t-1}} \sqrt{q_t^*}^{\lambda_{u+1} \oplus \dots \oplus \lambda_{t-1}} = \\ &= (-1)^{\sum_{i=1}^{t-1} \lambda_i \mu_i} \sqrt{(q_1^*)^{\lambda_1} \dots (q_{t-1}^*)^{\lambda_{t-1}} (q_t^*)^{\lambda_{u+1} \oplus \dots \oplus \lambda_{t-1}}}, \end{aligned}$$

то $\tau_\mu|_{\mathbb{Q}(g_\lambda)}$ действует тождественно при $\sum_{i=1}^{t-1} \lambda_i \mu_i \equiv 0 \pmod{2}$ и совпадает с σ_λ в противном случае.

Обозначим $\max_{\mu \neq (0, \dots, 0)} z_{\mu, n_\mu} = \varepsilon$. Тогда по лемме (6) для всех $\lambda \neq (0, \dots, 0)$ справедлива двусторонняя оценка $\frac{\varepsilon}{\sqrt{|d|}} \leq z_{\lambda, n_\lambda} \leq \varepsilon$. Из (29), (30) и леммы 5 следует, что $1 \leq z_{\lambda, n_\lambda} |\sigma_\lambda(z_{\lambda, n_\lambda})| < \sqrt{\delta_\lambda} \leq \sqrt{|d|}$. Следовательно, $\frac{1}{\varepsilon} \leq |\sigma_\lambda(z_{\lambda, n_\lambda})| \leq \frac{|d|}{\varepsilon}$. По построению $Z = \prod_{\lambda \neq 0} |\sigma_\lambda(z_{\lambda, n_\lambda})| \leq \left(\frac{|d|}{\varepsilon}\right)^{m-1}$, откуда $\varepsilon \leq \frac{|d|}{Z^{\frac{1}{m-1}}}$.

Пусть $\lambda \neq 0$. Тогда условие $\sum_{i=1}^{t-1} \lambda_i \mu_i \equiv 0 \pmod{2}$, рассматриваемое как уравнение на $\mu \in \{0, 1\}^{t-1}$, имеет ровно $n/2$ решений, одно из которых нулевое. Получим

$$\left| \tau_\lambda \left(\sum_{\mu} A_\mu \omega_\mu^* \right) \right| = \prod_{\substack{2|\sum_i \lambda_i \mu_i \\ \mu \neq 0}} |\sigma_\mu(z_{\mu, n_\mu})| \cdot \prod_{2 \nmid \sum_i \lambda_i \mu_i} |z_{\mu, n_\mu}| \leq \left(\frac{|d|}{\varepsilon}\right)^{\frac{m}{2}-1} \varepsilon^{\frac{m}{2}} = |d|^{\frac{m}{2}-1} \varepsilon \leq \frac{\sqrt{|d|}^m}{Z^{\frac{1}{m-1}}}.$$

Остаётся доказать, что алгоритм завершает работу за $O(\ln N_0)$ итераций. Доказанная часть теоремы позволяет применить теорему 15, согласно которой в процессе работы алгоритма всегда выполнено неравенство $|A_{0, \dots, 0}| \geq |\mathfrak{M}(Id)\omega_{0, \dots, 0}|Z - C\sqrt{|d|}^m$, где константы $\mathfrak{M}(Id)\omega_{0, \dots, 0} \neq 0$ и $C\sqrt{|d|}^m$ зависят только от базисов.

В силу (30) $Z = \prod_{\lambda \neq 0} \frac{y'_{\lambda, n_\lambda}}{z_{\lambda, n_\lambda}} \geq \left(\prod_{\lambda \neq 0} z_{\lambda, n_\lambda} \right)^{-1}$, откуда в силу (29), (25) и (26)

$$Z \geq \left(\prod_{\lambda \neq 0} |P_{\lambda, n_\lambda-1} - Q_{\lambda, n_\lambda-1} g_\lambda| \right)^{-1} \geq \prod_{\lambda \neq 0} Q_{\lambda, n_\lambda} \geq \prod_{\lambda \neq 0} 2^{\frac{n_\lambda-1}{2}} = 2^{\frac{\sum_{\lambda \neq 0} n_\lambda - (m-1)}{2}}.$$

Сумма $\sum_{\lambda \neq 0} n_\lambda$ есть общее число итераций алгоритма. Таким образом, после $O(\ln N_0)$

итераций алгоритма достигается неравенство $Z \geq \frac{N_0 + C\sqrt{|d|^m}}{|\mathfrak{M}(Id)\omega_{0,\dots,0}|}$, из которого следует, что $|A_{0,\dots,0}| \geq N_0$, что и требовалось доказать. ■

7. Вычисление элемента поля родов по приближённому значению

Восстановим числа $b_\mu \in \mathbb{Z}$ по приближённому значению числа $\sum_{\mu} b_\mu \beta_\mu$, а также числа $b'_\mu \in \mathbb{Z}$ — по приближённому значению $\sum_{\mu} b'_\mu \beta_\mu^*$. В п. 5 получены априорные оценки

$$\left| \tau_\lambda \left(\sum_{\mu} b_\mu \beta_\mu \right) \right| \leq T_0, \quad \left| \tau_\lambda \left(\sum_{\mu} b'_\mu \beta_\mu^* \right) \right| \leq T_0, \quad (32)$$

где T_0 — некоторое выражение, зависящее только от D . В п. 6 построен набор совместных приближений к числам $\beta_\mu/\beta_{0,\dots,0}$ и $\beta_\mu^*/\beta_{0,\dots,0}^*$ с точностью, зависящей от параметра N_0 .

Построенные приближения удовлетворяют теореме 16, которую и будем использовать. (Можно доказать, что любые совместные приближения Λ_i к базису W_i , удовлетворяющие оценке вида $\left| \frac{\Lambda_i}{\Lambda_1} - \frac{W_i}{W_1} \right| \leq \frac{C'_i}{|\Lambda_1|^{1+\alpha}}$, также удовлетворяют последней оценке из теоремы 16 с показателем α вместо $1/(m-1)$ и некоторой константой в числителе.)

Продолжим использовать базисы ω_μ , ω_μ^* и функцию \mathfrak{M} , определённые в (23) для задачи поиска b_μ и в (24) для задачи поиска b'_μ . Легко видеть, что при таком определении помимо свойств 1–3 \mathfrak{M} -пар выполнено следующее свойство:

2'. Если $x \in \mathcal{O}_M$, то $\omega_\xi x$ разлагается по базису $\{\omega_\mu\}$ с целыми коэффициентами.

Для определённости будем работать с числами b_μ ; нахождение чисел b'_μ полностью аналогично.

Пусть $X_\eta \in \mathcal{O}_M$ — линейно независимый над \mathbb{Q} набор из $m = 2^{t-1}$ чисел. Например, можно взять $X_\eta = \beta_\eta$; или $X_{0,\dots,0} = 1$ и $X_\eta = g_\eta$ при $\eta \neq (0, \dots, 0)$, где g_η определены в предыдущем пункте. В силу условия 2'

$$\omega_\xi X_\eta = \sum_{\mu} x_{\mu\xi\eta} \omega_\mu, \quad (33)$$

где $x_{\mu\xi\eta} \in \mathbb{Z}$. (Выбор $X_\eta = g_\eta$ удобен тем, что коэффициенты $x_{\mu\xi\eta}$ совпадают с рассмотренными в предыдущем пункте коэффициентами $c_{\mu\xi\eta}$ с перестановкой β_μ и β_μ^* . Выбор $X_\eta = \beta_\eta$ приводит к несколько меньшим по абсолютной величине числам.)

Зададимся точностью $\varepsilon > 0$ и вычислим $\sum_{\xi} b_\xi \beta_\xi$ с точностью ε , т. е. найдём число γ ,

такое, что $\left| \sum_{\xi} b_\xi \beta_\xi - \gamma \right| < \varepsilon$. Разделив это неравенство на $\beta_{0,\dots,0}$ и умножив на X_η , получим

$$\left| \sum_{\xi} b_\xi \omega_\xi X_\eta - \frac{\gamma X_\eta}{\beta_{0,\dots,0}} \right| \leq \frac{\varepsilon |X_\eta|}{|\beta_{0,\dots,0}|},$$

$$\left| \sum_{\mu} \left(\sum_{\xi} b_\xi x_{\mu\xi\eta} \right) \omega_\mu - \frac{\gamma X_\eta}{\beta_{0,\dots,0}} \right| \leq \frac{\varepsilon |X_\eta|}{|\beta_{0,\dots,0}|}. \quad (34)$$

Обозначим $B_{\mu\eta} = \sum_{\xi} b_{\xi} x_{\mu\xi\eta} \in \mathbb{Z}$. Тогда для произвольного μ' из (22) следует, что

$$B_{\mu'\eta} = \sum_{\mu} B_{\mu\eta} \sum_{\lambda} \mathfrak{M}(\tau_{\lambda}) \tau_{\lambda}(\omega_{\mu} \omega_{\mu'}^*) = \sum_{\lambda} \mathfrak{M}(\tau_{\lambda}) \tau_{\lambda}(\omega_{\mu'}^*) \tau_{\lambda} \left(\sum_{\mu} B_{\mu\eta} \omega_{\mu} \right),$$

$$\sum_{\mu'} A_{\mu'} B_{\mu'\eta} = \sum_{\lambda} \mathfrak{M}(\tau_{\lambda}) \tau_{\lambda} \left(\sum_{\mu'} A_{\mu'} \omega_{\mu'}^* \right) \tau_{\lambda} \left(\sum_{\mu} B_{\mu\eta} \omega_{\mu} \right). \quad (35)$$

Слагаемое с $\lambda = 0$ будем рассматривать отдельно; в этом случае (34) даёт приближённое значение последнего сомножителя и оценку погрешности приближения. При $\lambda \neq 0$ оценку второго сомножителя даёт теорема 16. Найдём оценку последнего сомножителя: $\tau_{\lambda} \left(\sum_{\mu} B_{\mu\eta} \omega_{\mu} \right) = \tau_{\lambda} \left(\sum_{\xi} b_{\xi} \sum_{\mu} x_{\mu\xi\eta} \omega_{\mu} \right) = \tau_{\lambda} \left(\sum_{\xi} b_{\xi} \omega_{\xi} \right) \tau_{\lambda}(X_{\eta})$, откуда и из (32) следует, что $\left| \tau_{\lambda} \left(\sum_{\mu} B_{\mu\eta} \omega_{\mu} \right) \right| \leq T_0 |\tau_{\lambda}(X_{\eta})|$. Таким образом, из (35), (34) и теоремы 16 следует, что

$$\left| \sum_{\mu'} A_{\mu'} B_{\mu'\eta} - \mathfrak{M}(Id) Z \frac{\gamma X_{\eta}}{\beta_{0,\dots,0}} \right| \leq |\mathfrak{M}(Id) Z| \frac{\varepsilon |X_{\eta}|}{|\beta_{0,\dots,0}|} + \sum_{\lambda \neq 0} |\mathfrak{M}(\tau_{\lambda})| \frac{\sqrt{|d|}^m}{Z^{\frac{1}{m-1}}} T_0 |\tau_{\lambda}(X_{\eta})|, \quad (36)$$

где, как и раньше, $Z = \sum_{\mu} A_{\mu} \omega_{\mu}^*$.

Второе слагаемое представляет из себя отношение некоторой константы и $Z^{\frac{1}{m-1}}$. Поскольку $A_{0,\dots,0} = \Lambda_1$, неравенство (21) показывает, что можно выбрать порог N_0 в алгоритме так, чтобы было выполнено неравенство

$$Z > \left(4 \sum_{\lambda \neq 0} |\mathfrak{M}(\tau_{\lambda}) \tau_{\lambda}(X_{\eta})| \sqrt{|d|}^m T_0 \right)^{m-1}, \quad (37)$$

а тогда второй член в правой части (36) меньше $1/4$.

Выбрав такой порог N_0 и построив совместные приближения A_{μ} , вычислим Z . Выберем ε так, чтобы для всех η было выполнено неравенство

$$\varepsilon < \frac{1}{4} \frac{|\beta_{0,\dots,0}|}{|\mathfrak{M}(Id) X_{\eta}| Z}. \quad (38)$$

Тогда первый член в правой части (36) также будет меньше $1/4$, а следовательно, левая часть (36) меньше $1/2$. Поскольку $\sum_{\mu'} A_{\mu'} B_{\mu'\eta} \in \mathbb{Z}$, то можно восстановить точное

значение этой суммы, округляя $\mathfrak{M}(Id) Z \frac{\gamma X_{\eta}}{\beta_{0,\dots,0}}$.

Возвращаясь к b_{ξ} , получаем на них систему линейных уравнений с левой частью

$$\sum_{\mu} A_{\mu} B_{\mu\eta} = \sum_{\xi} \left(\sum_{\mu} A_{\mu} x_{\mu\xi\eta} \right) b_{\xi}. \quad (39)$$

Лемма 7. Матрица $\left(\sum_{\mu} A_{\mu} x_{\mu\xi\eta} \right)_{\xi, \eta \in \{0,1\}^{t-1}}$ невырождена.

Доказательство. Предположим, что матрица вырождена. Это эквивалентно существованию таких чисел $y_\eta \in \mathbb{Q}$, не равных одновременно нулю, что

$$\sum_{\eta} \sum_{\mu} A_{\mu} x_{\mu\xi\eta} y_{\eta} = 0. \quad (40)$$

Пусть η фиксировано. Введём квадратные матрицы с элементами $(M_1)_{\mu_1\mu_2} = \tau_{\mu_1}(\omega_{\mu_2})$, $(M_2)_{\mu'_1\mu'_2} = \tau_{\mu'_1}(\omega_{\mu'_2}^*)$, $(X)_{\mu''_1\mu''_2} = x_{\mu''_1\mu''_2\eta}$ и квадратные диагональные матрицы M_3 с элементами $\mathfrak{M}(\tau_{\mu})$, M_4 с элементами $\tau_{\mu}(X_{\eta})$. Равенство (22) можно интерпретировать как матричное равенство $M_1^T M_3 M_2 = E$, где E — единичная матрица. В частности, M_1 , M_2 и M_3 обратимы. Совокупность всех равенств, получающихся из (33) применением различных τ_{μ} , можно интерпретировать как матричное равенство $M_4 M_1 = M_1 X$. Отсюда $X = M_1^{-1} M_4 M_1$, $X^T = M_1^T M_4 (M_1^T)^{-1} = M_2^{-1} M_3^{-1} M_4 M_3 M_2$. Поскольку любые диагональные матрицы, в том числе M_3 и M_4 , коммутируют, то $M_2 X^T = M_4 M_2$. Сравнивая элемент на пересечении первой строки и столбца μ , находим $\sum_{\xi} \omega_{\xi}^* x_{\mu\xi\eta} = X_{\eta} \omega_{\mu}^*$.

Далее η не фиксировано. Умножим (40) на ω_{ξ}^* и сложим по всем $\xi \in \{0, 1\}^{t-1}$:

$$\sum_{\eta} \sum_{\mu} A_{\mu} X_{\eta} \omega_{\mu}^* y_{\eta} = 0, \quad \left(\sum_{\eta} X_{\eta} y_{\eta} \right) \left(\sum_{\mu} A_{\mu} \omega_{\mu}^* \right) = 0.$$

Но первый множитель не равен нулю, поскольку X_{η} выбирались линейно независимыми над \mathbb{Q} , а $y_{\eta} \in \mathbb{Q}$ не все равны нулю. Второй множитель не равен нулю по теореме 16. Полученное противоречие доказывает лемму 7. ■

Таким образом, для нахождения $\{b_{\mu}\}$ остаётся решить заведомо невырожденную систему размером $t \times t$, например стандартным методом Гаусса.

Общая схема предлагаемой модификации метода комплексного умножения такова.

- 1) Выберем числа $q = p^n$, $\hat{u}, \hat{v}, D \in \mathbb{Z}$ как в шаге 1 базового алгоритма из п. 1.2. Кривая, которую мы построим, будет определена над \mathbb{F}_q и будет иметь порядок $q + 1 - \hat{u}$.
- 2) Вычислим все примитивные формы, оценку T_0 по формуле (18), порог N_0 , такой, чтобы была выполнена оценка (37), с использованием формулы (21). Применим алгоритм нахождения совместных приближений из п. 6.
- 3) Вычислим необходимую точность ε по формуле (38). Найдём многочлен $\hat{H}_D[j]$ по определению (16) приближённо с точностью ε .
- 4) Для каждого коэффициента найденного многочлена определим разложение его удвоенной вещественной части по базису β_{μ} , составив систему линейных уравнений с левой частью (39) с использованием (36) и решив эту систему. Аналогично определим разложение удвоенной мнимой части каждого коэффициента по базису β_{μ}^* . (Если известно, что коэффициент вещественный, то не нужно вычислять разложение мнимой части и не нужно умножать вещественную часть на 2.)
- 5) Редуцируем многочлен по модулю любого простого идеала, лежащего над p , в кольце \mathcal{O}_{K_G} ; получим некоторый многочлен над \mathbb{F}_q , который разлагается на линейные множители над \mathbb{F}_q . Вычислим какой-нибудь из его корней и построим эллиптическую кривую E'' над \mathbb{F}_q с j -инвариантом, равным вычисленному корню.
- 6) Если порядок кривой E'' оказался не таким, какой требуется, применим изоморфизм, описанный в п. 1.2.

Как и в исходном методе, в описанной модификации вместо модулярного инварианта j можно использовать другие функции, описанные в п. 1.3. Для этого следует скорректировать оценку T_0 на шаге 2, как описано в п. 5, на шаге 3 вычислить многочлен $\hat{H}_D[\theta, \alpha_*]$, а на шаге 5 после нахождения корня редуцированного многочлена вычислить j -инвариант как функцию от найденного корня, описанную в п. 1.3.

Автор благодарит своего научного руководителя Михаила Алексеевича Черепнёва за постановку задачи, продуктивные обсуждения и полезные замечания, позволившие уточнить и прояснить текст.

ЛИТЕРАТУРА

1. *Koblitz N.* Elliptic curve cryptosystems // *Math. Comput.* 1987. V. 48. P. 203–209.
2. *Miller V. S.* Uses of elliptic curves in cryptography // *LNCS.* 1986. V. 218. P. 417–426.
3. *Lenstra H. W.* Factoring integers with elliptic curves // *Ann. Math.* 1987. V. 126. P. 649–673.
4. *Atkin A. O. L. and Morain F.* Elliptic curves and primality proving // *Math. Comput.* 1993. V. 61. No. 203. P. 29–68.
5. *Cox D. A.* Primes of the form $x^2 + ny^2$. New York: Wiley, 1989.
6. *Weber H.* Lehrbuch der Algebra. 3rd edition. New York: Chelsea Publishing Company, 1908. V. 3.
7. *Silverman J. H.* The Arithmetic of Elliptic Curves. Springer, 1986.
8. *Ленг С.* Эллиптические функции. М.: Наука, 1984.
9. *Айерленд К., Роузен М.* Классическое введение в современную теорию чисел. М.: Мир, 1987.
10. *Cornacchia G.* Su di un metodo per la risoluzione in numeri interi dell' equazione $\sum_{h=0}^n C_h x^{n-h} y^h = P$ // *Giorn. Mat. Batt.* 1908. No. 46. P. 33–90.
11. *Baier H.* Efficient algorithms for generating elliptic curves over finite fields suitable for use in cryptography. Department of Computer Science, Technical University of Darmstadt, 2002.
12. *Konstantinou E., Kontogeorgis A., Stamatiou Y. C., and Zaroliagis C. D.* On the Efficient Generation of Prime-Order Elliptic Curves // *J. Cryptol.* 2010. V. 23. No. 3. P. 477–503.
13. *Deuring M.* Die Typen der Multiplikatorenringe elliptischer Funktionenkörper // *Abh. Math. Sem. Hansischen Univ.* 1941. B. 14. S. 197–272.
14. *Lay G.-J. and Zimmer H. G.* Constructing elliptic curves with given group order over large finite fields // *LNCS.* 1994. V. 877. P. 250–263.
15. *Von Schrutka L.* Ein Beweis für die Zerlegbarkeit der Primzahlen von der Form $6n + 1$ in ein einfaches und ein dreifaches Quadrat // *J. Reine Ang. Math.* 1911. B. 140. S. 252–265.
16. *Jacobstahl E.* Über die Darstellung der Primzahlen der Form $4n + 1$ als Summe zweier Quadrate // *J. Reine Ang. Math.* 1907. B. 132. S. 238–245.
17. *Schertz R.* Weber's class invariants revisited // *J. Théor. Nomb. Bord.* 2002. V. 14. No. 1. P. 325–343.
18. *Enge A. and Schertz R.* Constructing elliptic curves over finite fields using double eta-quotients // *J. Théor. Nomb. Bord.* 2004. V. 16. No. 3. P. 555–568.
19. *Cohn H.* Introduction to the construction of class fields. Cambridge University Press, 1985.
20. *Ленг С.* Алгебраические числа. М.: Мир, 1966.
21. *Enge A.* The complexity of class polynomial computation via floating point approximations // *Math. Comput.* 2009. V. 78. No. 266. P. 1089–1107.
22. *Brisebarre N. and Philibert G.* Effective lower and upper bounds for the Fourier coefficients of powers of the modular invariant j // *J. Raman. Math. Soc.* 2005. V. 20. P. 255–282.

23. *Enge A. and Morain F.* Comparing invariants for class fields of imaginary quadratic fields // LNCS. 2002. V. 2369. P. 252–266.
24. *Brentjes A. J.* Multi-dimensional continued fraction algorithms. Amsterdam: Mathematisch Centrum, 1981.
25. *Peck L. G.* Simultaneous rational approximations to algebraic numbers // Bull. Amer. Math. Soc. 1961. V. 67. P. 197–201.
26. *Хинчин А. Я.* Цепные дроби. М.: Наука, 1978.
27. *Венков Б. А.* Элементарная теория чисел. ОНТИ НКТП СССР, 1937.
28. *Гречников Е. А.* Оптимизация метода с комплексным умножением построения эллиптической кривой // Деп. в ВИНТИ 21.06.11, №305-В2011. М.: МГУ им. М. В. Ломоносова, 2011.

МАТЕМАТИЧЕСКИЕ ОСНОВЫ КОМПЬЮТЕРНОЙ БЕЗОПАСНОСТИ

DOI 10.17223/20710410/13/4

УДК 004.4'2+004.43

ВНЕДРЕНИЕ ПОЛИТИК БЕЗОПАСНОСТИ В ПРОГРАММНЫЕ СИСТЕМЫ ОБРАБОТКИ ИНФОРМАЦИИ

Д. А. Стефанцов

*Томский государственный университет, г. Томск, Россия***E-mail:** d.a.stefantsov@isc.tsu.ru

Рассматривается проблема защиты систем обработки информации (СОИ) посредством интеграции их с политиками безопасности (ПБ). Проанализированы существующие методы решения этой проблемы, отмечены их недостатки и предложен оригинальный метод её решения с помощью аспектно-ориентированного программирования (АОП), лишённый этих недостатков. В отличие от традиционных реализаций АОП, в данном методе аспект ПБ присоединяется к СОИ посредством соединительного модуля без изменения программных модулей СОИ и ПБ, написанных независимо друг от друга и от соединительного модуля. Для реализации этого метода созданы инструментальные средства в составе языка АОП AspectTalk, виртуальной машины и транслятора с языка AspectTalk в язык виртуальной машины. Работа содержит краткое описание предложенного метода и перечисленных инструментальных средств его реализации.

Ключевые слова: *система обработки информации, политика безопасности, аспектно-ориентированное программирование, AspectTalk, виртуальная машина.*

Введение

Защита информации, хранимой и преобразуемой в системах обработки информации (СОИ), является актуальной проблемой с момента появления многопользовательских компьютерных систем [1]. При разработке защиты таких систем определяется модель нарушителя в виде формального описания набора угроз системе и/или атак на неё, а также политика безопасности (ПБ) в виде набора формальных правил противодействия последним [2]. Невозможность компрометации системы при условии следования правилам ПБ доказывается соответствующими теоремами безопасности [1–3].

Одной из первых формальных ПБ, разработанных для реализации в вычислительных системах, является модель Белла — ЛаПадулы [1]. Обзор большинства существующих моделей безопасности можно прочитать в [3]. Министерством обороны США представлена классификация вычислительных систем на основе реализации ПБ для определённых моделей нарушителя [4].

Помимо политик разграничения доступа, в понятие ПБ будем включать любые требования к защите СОИ, например использование криптографических средств защиты информации и специальных методов преобразования информации, таких, как фильтрация и кодирование.

Программную составляющую защищённой СОИ можно разделить на две части (подсистемы) — часть, реализующую целевую обработку информации (далее — часть

ОИ), и часть, реализующую программную модель ПБ (далее — часть ПБ). Примерами подобной модели могут служить следующие алгоритмы и структуры данных, реализующие действия, предписываемые ПБ:

- 1) учётные записи пользователей, списки прав доступа — при реализации политики разграничения доступа;
- 2) алгоритмы шифрования и цифровой подписи, криптографические протоколы, хранилища ключевой информации — при реализации криптографической защиты данных;
- 3) алгоритмы фильтрации вводимых данных для предотвращения SQL-инъекций и XSS-атак.

Для соединения этих частей в одну программу часть ОИ обычно изменяется таким образом, что при совершении действий по обработке информации производится обращение к части ПБ для определения возможности доступа к запрашиваемой информации или для её преобразования. Эти изменения делают текст программы части ОИ зависимым от текста программы части ПБ.

При изменении модели нарушителя соответствующие изменения вносятся в часть ПБ, что может повлечь за собой необходимость изменения части ОИ. Примером подобных изменений может служить реализация политики мандатного разграничения доступа SELinux для операционной системы (ОС) GNU/Linux, ранее обладавшей только дискреционной политикой разграничения доступа [5]. Тесная интеграция программных реализаций СОИ и ПБ является препятствием к внесению изменений в ПБ.

Примером СОИ, в которой части ОИ и ПБ реализованы отдельно, может служить ОС Mac OS X 10.4. В ней при необходимости организации доступа субъекта системы к её объекту вызывается специальная функция подсистемы `kauth` [6]. Эта функция опрашивает множество специальных модулей, ответственных за реализацию политики разграничения доступа. На основании ответов, полученных от модулей, функция вычисляет ответ подсистемы безопасности на запрос доступа: разрешение или отказ. Модули, выносящие решение в соответствии с некоторой ПБ, разрабатываются и реализуются независимо от ядра ОС в соответствии со специальными правилами и могут быть загружены в память ОС администратором системы. Таким образом, для реализации некоторой политики разграничения доступа с помощью подсистемы `kauth` необходимо описать совокупность модулей на некотором языке программирования, скомпилировать их и загрузить в память ОС. На рис. 1 схематически изображена работа ОС с реализованной в ней подсистемой `kauth`.

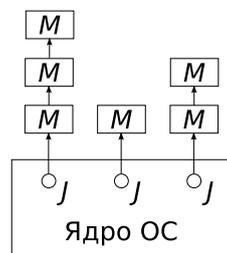


Рис. 1. Подсистема `kauth` ОС Mac OS X 10.4: цепочки модулей `M` определяют возможность предоставления доступов `J`

Подсистема `kauth` реализована также в ОС NetBSD [7]. В работе [8] показана возможность реализации механизма `jail` ОС FreeBSD в ОС NetBSD средствами подсистемы `kauth`.

К недостаткам подсистемы `kauth` можно отнести невозможность реализации ПБ, не являющейся политикой разграничения доступа.

Автором разработана технология и инструментальная среда создания защищённых СОИ в виде совокупности независимых частей ОИ и ПБ произвольного вида, соединяемых способом, исключающим необходимость их изменения. В основе этого способа лежит аспектно-ориентированное программирование (АОП) [9], модифицированное для случая, когда одна из объединяемых подсистем реализует ПБ, а именно: интеграция частей ОИ и ПБ осуществляется при помощи простых соединительных модулей, которые зависят одновременно от текста программы ОИ и текста присоединяемого аспекта ПБ. Части ОИ и ПБ и соединительные модули представляются в виде исходных текстов на языке АОП AspectTalk [10], специально разработанном для этой цели. В нём исходные тексты частей ОИ и ПБ являются описаниями классов объектов их предметных областей, а соединительные модули устанавливают соответствие между классами этих частей. В процессе трансляции соответствующие классы объединяются в один. Трансляция выполняется в язык интерпретируемой виртуальной машины (ВМ).

Краткое сообщение об этих средствах создания защищённых СОИ можно найти в [11]. Более развёрнутое изложение основных элементов данной технологии является целью настоящей работы.

1. Краткая характеристика АОП

АОП — это способ программирования, при котором главная подсистема (часть, выполняющая основную функцию системы) может быть реализована независимо от подчинённой подсистемы (части, выполняющей дополнительную функцию — например, реализацию ПБ) [10].

Рассмотрим СОИ, реализованную с помощью традиционного процедурного подхода и состоящую из программы P и библиотеки L (рис. 2). Если в точках J программы P необходим вызов процедуры F библиотеки L , он будет осуществлён явно, что сделает текст программы P зависимым от текста библиотеки L . В случае замены или удаления библиотеки L необходимо изменение текста программы P .

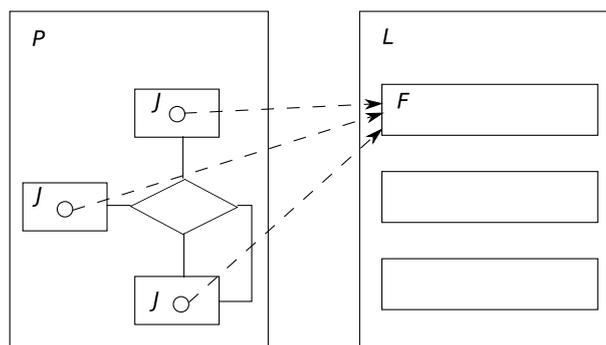


Рис. 2. Использование традиционного подхода в программировании. Текст программы P зависит от текста библиотеки L

Разработка программ с помощью АОП основана на использовании неявных вызовов процедур. Рассмотрим СОИ, состоящую из программы P и аспекта A (рис. 3). Аспект — это библиотека специального вида, состоящая из описаний состояний программы, а также специальных алгоритмов. Между описаниями и алгоритмами устанавливается соответствие: всякий раз, когда программа достигает состояния, подхо-

дящего под описание C , запускается соответствующий этому описанию алгоритм D . При этом J — это точки выполнения программы, в которых достигаются состояния, подходящие под описание C , и осуществляется неявный вызов алгоритма D . В данном случае текст аспекта A зависит от текста программы P , к которой этот аспект применяется, но текст программы P не зависит от текста аспекта A . К недостаткам данного варианта АОП можно отнести зависимость аспектов от программы, к которой они применяются, что затрудняет перенос аспектов в другие СОИ [12].

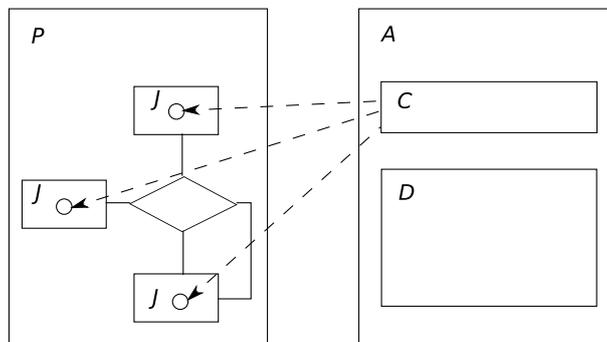


Рис. 3. Использование АОП. Текст аспекта A зависит от текста программы P

АОП не является самостоятельным способом программирования, но реализуется в виде расширения некоторого традиционного способа программирования [12]. А именно, программа P и алгоритм D могут быть реализованы без применения аспектов. Аспектное расширение традиционного подхода — это присутствие описаний C состояний программы P .

2. Метод защиты СОИ с помощью АОП

Опишем метод создания защищённых СОИ средствами АОП, в котором тексты программы и аспектов разрабатываются независимо друг от друга, а для интеграции частей используются специальные соединительные модули, которые зависят одновременно от текста программы и текста присоединяемого аспекта (рис. 4). Соединительные модули предполагаются проще соединяемых ими частей СОИ.

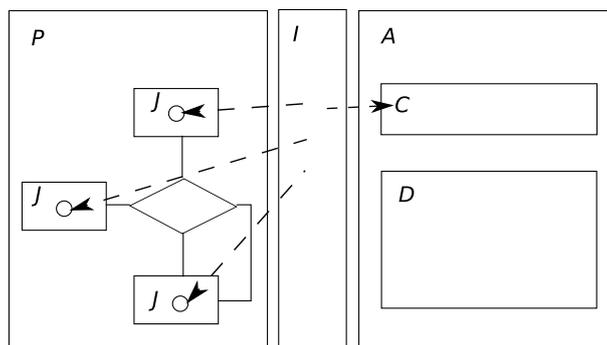


Рис. 4. СОИ получена интеграцией частей с помощью соединительных модулей

Части ОИ и ПБ предполагаются представленными в виде исходных текстов на объектно-ориентированном языке программирования. Предметной областью части ОИ является целевая предметная область СОИ, предметной областью части ПБ является политика безопасности. Исходные тексты частей ОИ и ПБ — это описания классов объектов соответствующих предметных областей.

Предлагаемая технология основана на установлении соответствия между классами объектов данных предметных областей. На рис. 5 показано соответствие между классами двух предметных областей: Unix-подобной ОС и дискреционной политики разграничения доступа.

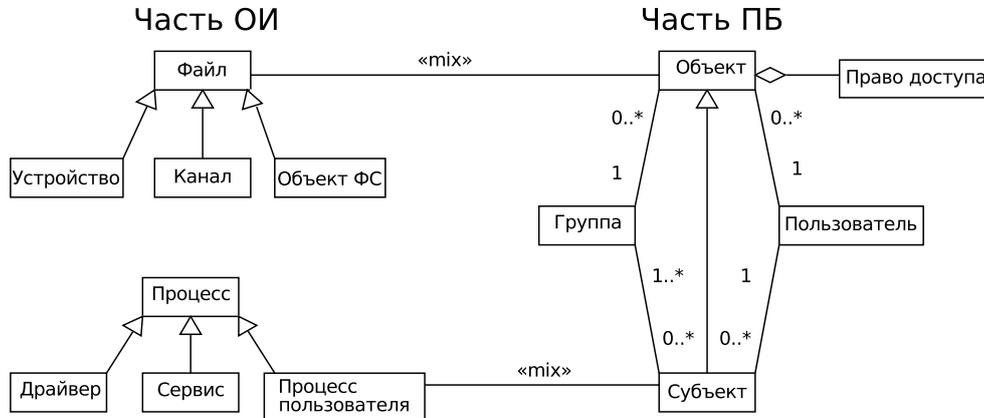


Рис. 5. Диаграмма классов и примесей Unix-подобной системы обработки файлов

Классы подчинённой системы (в данном случае — ПБ) называются примесями, а соответствие примесей части ПБ и классов части ОИ обозначается «mix». По данному соответствию на этапе трансляции происходит объединение классов и примесей в составной класс, множество член-данных которого — это объединение множеств член-данных соответствующих класса и примесей, а множество член-функций — объединение множеств член-функций соответствующих класса и примесей (рис. 6). Соответствие «mix» является альтернативной реализацией механизма множественного наследования.



Рис. 6. Объединение класса и примеси на этапе трансляции

Для примесей определяются также специальные алгоритмы, называемые конвертами. При указании соответствия между классами и примесями для конвертов задаются шаблоны строк в виде регулярных выражений. Всякий раз, когда будет вызвана член-функция с именем α некоторого объекта из части ОИ, α будет проверено на соответствие шаблонам для конвертов соответствующих примесей. При совпадении вместо вызываемого алгоритма будет запущен соответствующий алгоритм-конверт. Алгоритм-конверт получает в качестве параметров объект, вызвавший исходную член-функцию, объект, член-функция которого была вызвана, имя член-функции, а также параметры вызова. Алгоритм-конверт формирует возвращаемое значение, в процессе чего может вызывать исходную член-функцию. Механизм конвертов позволяет реализовать вход программы в аспект.

Рассмотренный механизм примесей похож на описанный в [12], однако в [12] примесь реализована в виде базовой конструкции языка. Предлагаемый же подход основан на реализации в языке протокола метаобъектов [13] в качестве базовой конструкции с последующей реализацией примесей на самом языке программирования. Это позволяет упростить семантическую модель языка и не различать примеси, модифицирующие алгоритмы, и примеси, модифицирующие структуры данных, как это сделано в [12].

Таким образом, реализован следующий метод построения СОИ, защищённых ПБ:

- 1) часть ОИ описывается в терминах классов;
- 2) часть ПБ описывается в терминах примесей;
- 3) соединительные модули описываются в виде сопоставлений класс — примесь, а также указания шаблонов для алгоритмов-конвертов.

3. Краткая характеристика языка AspectTalk

Для создания СОИ, защищённых с помощью АОП, необходим специальный аспектно-ориентированный язык программирования (АОЯП), на котором описываются программа и её аспекты. Первым АОЯП принято считать AspectJ [14, 15] — расширение объектно-ориентированного языка программирования Java. В AspectJ описания S состояний программы P (см. рис. 3) даются в виде набора синтаксических конструкций, и состояние считается достигнутым, если выполняется действие, соответствующее той или иной конструкции языка. В других АОЯП, например в MetaclassTalk [12], под состоянием понимается состояние специальных переменных. О других способах реализации АОП можно прочитать в [16–18].

Для экспериментального исследования изложенного выше метода создания защищённых СОИ разработан и реализован АОЯП AspectTalk [10]. За его основу взят диалект Little Smalltalk объектно-ориентированного языка программирования Smalltalk [19].

Все типы данных в AspectTalk являются объектами. Среди объектов выделяется подмножество, называемое классами. С помощью классов и иерархии их наследования реализуются механизмы создания объектов и установления соответствия между объектами и их член-данными и член-функциями.

В языке AspectTalk декларативные конструкции языка Smalltalk, отвечающие, в том числе, за объявление классов и анализируемые на этапе трансляции, заменены на совокупности элементарных операций, которые исполняются на этапе выполнения программы. Это позволяет уменьшить число конструкций языка.

Язык AspectTalk содержит следующие базовые операции:

- 1) примитивы VM, с помощью которых производятся низкоуровневые действия, такие, как сложение чисел, работа с файлами;
- 2) посылка сообщения — это базовая операция, которая приводит к вызову член-функции, определяемой в соответствии с иерархией наследования классов;
- 3) возврат результата — операция, приводящая к завершению работы член-функции и возврату выполнения программы на следующую команду после соответствующей посылки сообщения;
- 4) присвоение — операция, с помощью которой переменным присваиваются указатели на объекты.

Как и в Smalltalk, в AspectTalk есть специальный тип данных — метаклассы, но, в отличие от Smalltalk, метаклассы в AspectTalk не только являются классами классов, но и позволяют программисту давать VM дополнительные указания о работе объектной системы. Эти указания определяются в виде обработчиков операции посылки сообщения от объекта к объекту и наследуются метаклассами в иерархии наследования метаклассов. Более подробное описание языка AspectTalk с примером реализации с его помощью некоторой ПБ можно найти в [10].

4. Краткая характеристика интерпретатора

В большинстве случаев АОЯП разделяется на две составляющие: базовый язык программирования, с помощью которого реализуются основная часть программы и алгоритмы аспектов, и аспектное расширение, с помощью которого аспекты присоединяются к программе. Транслятор с такого языка реализуется, в основном, следующим образом. Этап трансляции разделяется на два шага. На первом шаге программа, написанная на базовом языке программирования совместно с его аспектным расширением, транслируется в программу, содержащую только конструкции базового языка программирования. В результате выполнения этого шага все неявные вызовы процедур аспекта, которые могут быть определены без запуска программы, заменяются на явные в автоматическом режиме. Если вызов процедуры может быть осуществлён на основе информации, доступной только во время выполнения программы, то во всех предположительных местах вызова помещается обращение к специальной библиотеке, которая определяет необходимость вызова процедуры аспекта. На втором шаге осуществляется трансляция с базового языка программирования на язык машины, выполняющей программу.

Возможен другой способ реализации транслятора с АОЯП, используемый в данной работе, — в виде VM, внутренние структуры данных и алгоритмы которой реализуют программную модель АОЯП, и транслятора с АОЯП в язык VM. В этом случае трансляция осуществляется за один шаг — программа, написанная на базовом языке программирования и его аспектном расширении, транслируется в язык VM напрямую. Совокупность транслятора и VM далее будем называть интерпретатором. Интерпретация менее эффективна по времени и по памяти, чем компиляция, однако реализуется более простым способом и более наглядна.

В интерпретаторе с языка AspectTalk во многом повторяется структура интерпретатора с языка Smalltalk [19]. Основные принципы, применяемые для перевода программы с AspectTalk в язык VM, не новы и описаны в [20].

VM интерпретатора с языка AspectTalk является стековой машиной, язык которой представляет собой польскую инверсную запись команд, с помощью которых производится манипуляция данными, а также обращение к ОС, в которой запускается интерпретатор. Основной тип данных VM — объект. Экземпляры этого типа данных

помещаются и извлекаются с вершины стека, а также преобразуются специальными командами.

Команды ВМ можно разделить на следующие группы:

- примитивы ВМ;
- команды создания объектов;
- команды, помещающие значение переменной на вершину стека;
- команды, сохраняющие в переменных объект с вершины стека;
- команды, посылающие сообщение объекту, находящемуся на вершине стека (при этом параметры сообщения также берутся из стека);
- команда возврата из процедуры;
- команды удаления и дублирования объекта на вершине стека.

Примитивы ВМ — это операции обращения к ВМ для выполнения низкоуровневых операций, такие, как сложение целых чисел, выделение памяти, вывод строки на экран, работа с файлами и т. д. Поиск переменных по имени осуществляется в соответствии с моделью лексического окружения, описанного в [21].

Основа работы интерпретатора — объектная модель. В этой модели вся система представлена в виде совокупности объектов, посылающих сообщения другим объектам. Получив сообщение, объект осуществляет его диспетчеризацию — поиск процедуры, которая должна быть выполнена в ответ на сообщение, после чего объект возвращает результат работы процедуры отправителю сообщения. Совокупность сообщений, для которых объект может найти соответствующую процедуру, называется протоколом этого объекта. В большинстве систем существует множество объектов, обладающих одинаковым протоколом. Для того чтобы упростить алгоритм поиска процедур для таких объектов, в объектно-ориентированном программировании вводится специальное понятие — класс. Класс — это объект, который осуществляет диспетчеризацию сообщений для объектов, обладающих одинаковым протоколом. Такие объекты называются экземплярами этого класса. Вводится также отношение наследования, которое организует классы в иерархию: потомок может обработать те же сообщения, что и предок, и ещё некоторые дополнительные сообщения.

В свою очередь, метаклассы — это классы классов. Они не только осуществляют диспетчеризацию сообщений, отправляемых классам, но и могут заменить у некоторых классов процедуру диспетчеризации сообщений, посылаемых экземплярам этих классов. Этот механизм называется протоколом метаобъектов и описан в [13]. С помощью этого механизма реализовано АОП в AspectTalk: всякий раз, когда некоторому объекту посылается сообщение, метакласс прерывает его диспетчеризацию и проверяет посылаемое сообщение на соответствие набору регулярных выражений. Если соответствие найдено, то запускается специальный алгоритм-конверт, соответствующий регулярному выражению. Если соответствия не найдено, то метакласс возобновляет диспетчеризацию сообщений через иерархию классов.

Следует отметить, что диспетчеризация сообщений, классы, метаклассы, протоколы метаобъектов и конверты не являются встроенными конструкциями языка ВМ, а описаны на самом AspectTalk в его библиотеке: для этого достаточно четырёх базовых операций языка.

Заключение

Разработаны технология и инструментальная среда создания защищённых СООИ в виде совокупности независимых частей (подсистем) — ОИ и ПБ произвольного вида,

соединённых простым способом, основанном на АОП, модифицированном для случая, когда одна из подсистем — это реализация ПБ. В составе разработанных средств:

- 1) язык программирования AspectTalk, обладающий конструкциями базового языка программирования Smalltalk и метаязыковыми конструкциями;
- 2) ВМ, выполняющая операции над данными, описанными на языке AspectTalk;
- 3) транслятор с языка AspectTalk в язык ВМ.

Разработанная технология предполагает:

- 1) описание ОИ в терминах классов;
- 2) описание ПБ в терминах примесей;
- 3) описание соединительных модулей путём сопоставлений класс — примесь и указания шаблонов для алгоритмов-конвертов.

Научная новизна исследования состоит в модификации аспектно-ориентированного подхода в программировании, состоящей в специализации средств метаязыка, а именно: в классическом АОП они используются при написании аспекта совместно с операциями объединения с основной программой, в модификации — только при написании соединительных модулей. Последнее упрощает процедуру повторного использования аспектов: вместо переписывания заново самого аспекта (в классическом подходе) переписывается только соединительный модуль (в новом подходе), который, как правило, много проще присоединяемого аспекта.

Использование разработанных методов и средств позволяет снизить затраты на внесение изменений в ПБ СОИ, а также повторно использовать реализации ПБ при разработке новых программных систем.

ЛИТЕРАТУРА

1. Bell D. E. and LaPadula L. J. Secure computer system: Unified exposition and multics interpretation: Tech. Rep. ESD-TR-75-306. The MITRE Corporation, 1976.
2. Landwehr C. E. Formal models for computer security // ACM Comput. Surv. 1981. V. 13. No. 3. P. 247–278.
3. Девянин П. Н. Анализ безопасности управления доступом и информационными потоками в компьютерных системах. М.: Радио и связь, 2006. 176 с.
4. DoD 5200.28-STD (Trusted Computer System Evaluation Criteria) USA: National Computer Security Center, 1985. 116 p.
5. <http://www.nsa.gov/research/selinux/index.shtml> — Security-Enhanced Linux. 2009.
6. http://developer.apple.com/library/mac/#technotes/tn2127/_index.html — Technical Note TN2127. Kernel Authorization. 2010.
7. <http://netbsd.gw.com/cgi-bin/man-cgi?kauth+9+NetBSD-current> — NetBSD Kernel Developer's Manual. kauth. 2009.
8. <http://2008.asiabsdcon.org/papers/P3A-paper.pdf> — Implementing Jails Under the kauth Framework. 2008.
9. Elrad T., Filman R. E., and Bader A. Aspect-Oriented Programming // Commun. ACM. 2001. V. 44. No. 10. P. 29–32.
10. Стефанцов Д. А. Реализация политик безопасности в компьютерных системах с помощью аспектно-ориентированного программирования // Прикладная дискретная математика. 2008. № 1. С. 94–100.
11. Стефанцов Д. А. Технология и инструментальная среда создания защищённых систем обработки информации // Прикладная дискретная математика. Приложение. 2009. № 1. С. 55–56.

12. *Bouraqadi N., Seriai A., and Leblanc G.* Towards unified aspect-oriented programming // ESUG 2005 Research Conference. Brussels, Belgium, 2005. 22 p.
13. *Kiczales G.* The Art of Meta-Object Protocol. The MIT Press, 1991. 345 p.
14. <http://eclipse.org/aspectj> — The AspectJ Project. 2011.
15. *Kiczales G., Hilsdale E., Hugunin J., et al.* Getting Started with AspectJ // Commun. ACM. 2001. V. 44. No. 10. P. 59–65.
16. *Diaz Pace J. A. and Campo M. R.* Analyzing the Role of Aspects in Software Design // Commun. ACM. 2001. V. 44. No. 10. P. 67–73.
17. *Lieberherr K., Orleans D., and Owingier J.* Aspect-Oriented Programming with Adaptive Methods // Commun. ACM. 2001. V. 44. No. 10. P. 39–41.
18. *Bergmans L. and Aksit M.* Composing Crosscutting Concerns Using Composition Filters // Commun. ACM. 2001. V. 44. No. 10. P. 51–57.
19. *Goldberg A. and Robson D.* Smalltalk 80 — The Language and its implementation. Addison-Wesley, 1983. V. 1. 714 p.
20. *Ахо А., Ульман Дж., Сети Р.* Компиляторы: принципы, технологии и инструменты. М.: Вильямс, 2003. 768 с.
21. *Абельсон Х., Сассман Дж.* Структура и интерпретация компьютерных программ. М.: Добросвет, 2006. 608 с.

ПРИКЛАДНАЯ ТЕОРИЯ ГРАФОВ

DOI 10.17223/20710410/13/5

УДК 519.178

ВЫЧИСЛИТЕЛЬНЫЕ АСПЕКТЫ ДРЕВОВИДНОЙ ШИРИНЫ ГРАФА

В. В. Быкова

*Институт математики Сибирского федерального университета, г. Красноярск, Россия***E-mail:** bykvalen@mail.ru

Дан краткий обзор современных результатов по проблеме вычисления древовидной ширины. Представлены и исследованы некоторые нижние и верхние оценки данного числового параметра графа. Предложены алгоритмические методы улучшения этих оценок.

Ключевые слова: алгоритмы на графах, частичные k -деревья, древовидная ширина.

Древовидная ширина — числовой параметр, характеризующий меру древовидности графа. Графы с ограниченной древовидной шириной образуют специальный класс графов, называемых частичными k -деревьями. Этот класс графов был введен четверть века назад Н. Робертсоном и П. Д. Сеймуром [1]. Широкий интерес к изучению древовидной ширины графа вызван тем, что многие NP-трудные задачи теории графов, возникающие в различных приложениях, в том числе при моделировании надежности и безопасности компьютерных и коммуникационных систем, полиномиально разрешимы, если модельный граф — частичное k -дерево. В этих условиях задача может быть решена методом динамического программирования, причем алгоритмическая эффективность достигается за счет разложения модельного графа на части с помощью небольших (мощности не более k) сепараторов [2]. С вычислительной точки зрения весьма сложно установить, имеет ли заданный произвольный граф ограниченную древовидную ширину. Доказана NP-полнота этой задачи [3, 4]. Лишь для некоторых классов графов древовидная ширина известна заранее или находится без особого труда. Например, древовидная ширина любого ациклического графа не превышает 1, а у всякого последовательно-параллельного графа древовидная ширина не более 2. Хордаловые графы — типичный пример класса графов, для которых задача определения древовидной ширины разрешима за полиномиальное время. Поскольку в общем случае данную задачу решить трудно, то актуальны границы возможных значений древовидной ширины графа и методы их улучшения.

1. Основные числовые параметры графа

Введем необходимые обозначения и определения. Везде далее $G = (V, E)$ — неориентированный конечный связный граф без петель и кратных ребер, $n = |V|$ и $m = |E|$. Обозначим через $G(A)$ подграф графа G , индуцированный множеством $A \subseteq V$. Говорят, что множество вершин $S \subseteq V$ разделяет несмежные вершины a и b графа G , если в графе $G(V \setminus S)$ вершины a и b принадлежат различным компонентам связности. Множество S при этом называется сепаратором, или (a, b) -сепаратором. Сепаратор минимальный, если он не содержит другой сепаратор в качестве собствен-

ного подмножества. Наименьший по размеру сепаратор графа G — наименьший сепаратор. Сепаратор размера 1 называется точкой сочленения графа. Вершина $a \in V$ считается симплициальной вершиной графа G , если ее окрестность $N(a)$ индуцирует в G клику. Множество вершин A называется кликой графа G , если $G(A)$ — полный граф, и максимальной кликой, если она не содержится в клике с большим количеством вершин. Размер наибольшей клики графа G обозначается $\varphi(G)$ и называется плотностью графа G . Хроматическое число графа G определяется как наименьшее возможное количество цветов $\chi(G)$, в которые можно раскрасить вершины графа так, что любые смежные вершины будут окрашены в разные цвета. Для произвольного графа G всегда $\varphi(G) \leq \chi(G)$, а для совершенного графа $\varphi(G) = \chi(G)$ [5]. Совершенными, например, являются полные, двудольные и хордальные графы. Граф G называется хордальным, если ни один из его индуцированных подграфов не является простым циклом длины $l \geq 4$. Всякий индуцированный подграф хордального графа также является хордальным. Известны полиномиальные по времени алгоритмы распознавания хордальности графа G , основанные на симплициальных вершинах и сепараторах [6, 7]. Нахождение чисел $\varphi(G)$, $\chi(G)$ для графа G общего вида — NP-трудные задачи [2]. Между тем для хордального графа G числа $\varphi(G)$, $\chi(G)$ могут быть вычислены за полиномиальное время [7]. Любой граф можно превратить в хордальный, добавив в него некоторое множество ребер. Триангуляцией графа $G = (V, E)$ называется хордальный граф $H = (V, E')$, который содержит G в качестве остова подграфа ($E \subseteq E'$). Триангуляция H — минимальная триангуляция графа G , если для G не существует другой триангуляции, которая является собственным подграфом графа H . Триангуляция, имеющая наименьшее число ребер, называется наименьшей. Число ребер наименьшей триангуляции является числовым параметром графа G и обозначается $\text{ch}(G)$. Задача нахождения $\text{ch}(G)$ для произвольного графа (ее также называют задачей наименьшего пополнения графа до хордального) NP-трудная. Заметим, что если граф G хордальный, то $\text{ch}(G) = m$. Числом вершинной связности $\kappa(G)$ графа G называется наименьшее число вершин, удаление которых приводит к несвязному или одновершинному графу. Граф G считается l -связным, если $\kappa(G) \geq l$. Согласно теореме Менгера [5], чтобы граф G был l -связен, необходимо и достаточно, чтобы любые две его несмежные вершины a и b были соединены не менее чем l вершинно-непересекающимися (a, b) -цепями. Другими словами, граф l -связен тогда и только тогда, когда каждый его (a, b) -сепаратор содержит по крайней мере l вершин. Значит, число $\kappa(G)$ определяет размер наименьшего сепаратора графа G (исключение составляет лишь полный n -вершинный граф K_n , в котором нет сепараторов вообще, хотя $\kappa(K_n) = n - 1$). Если $\delta(G)$ — минимальная степень вершин графа G , то всегда $\kappa(G) \leq \delta(G)$. Примечательно, что задача вычисления числа $\kappa(G)$ для графа G сводится к задаче о максимальном потоке и минимальном разрезе и разрешима за полиномиальное время [8].

Перечисленные выше числовые параметры графа давно известны и хорошо изучены в теории графов. Все они с различных позиций характеризуют структуру графа и являются его инвариантами в аспекте изоморфизма. Дополняет этот набор инвариантов древовидная ширина как мера близости графа к дереву.

2. Дерево декомпозиции и древовидная ширина графа

Древовидная ширина графа вычисляется через специальную графовую структуру, которая называется деревом декомпозиции. Дерево декомпозиции графа $G = (V, E)$ представляет собой пару (\mathcal{X}, T) , где $\mathcal{X} = \{X_i : i \in I\}$ — семейство подмножеств мно-

жества V , называемых «мешками»; $T=(I, W)$ — дерево, узлам которого сопоставлены эти «мешки», и выполняются следующие условия [1, 9]:

- 1) $\bigcup_{i \in I} X_i = V$;
- 2) для всякого ребра графа G обязательно имеется хотя бы один «мешок», содержащий обе вершины этого ребра;
- 3) для любой вершины $v \in V$ графа G множество узлов $\{i \in I : v \in X_i\}$ индуцирует связный подграф, являющийся поддеревом дерева T .

Ширина дерева декомпозиции (\mathcal{X}, T) равна $\max_{i \in I} \{|X_i| - 1\}$. Древовидная ширина (*treewidth*) графа G определяется как наименьшая ширина всех допустимых его деревьев декомпозиции и обозначается через $\text{tw}(G)$. Дерево декомпозиции (\mathcal{X}, T) ширины $\text{tw}(G)$ называется оптимальным деревом декомпозиции графа G . Заметим, что если рассматривать только деревья декомпозиции без кратных и вложенных «мешков», то для каждого связного графа G множество возможных деревьев декомпозиции непусто и конечно. Например, всегда существует тривиальное дерево декомпозиции, которое состоит из одного «мешка», содержащего все вершины графа. Однако такое дерево декомпозиции в большинстве случаев не является оптимальным. Когда граф G несвязен, то полагают $\text{tw}(G) = 0$. Следовательно, числовой параметр $\text{tw}(G)$ может быть вычислен для любого конечного графа G .

Дерево декомпозиции графа — это весьма полезная графовая структура, которая определяет не только древовидную ширину графа, но и отражает информацию о его кликах и сепараторах. Это подтверждает

Утверждение 1 [10]. Пусть задано некоторое дерево декомпозиции (\mathcal{X}, T) графа $G = (V, E)$. Тогда справедливы следующие высказывания:

- 1) если множество вершин $A \subseteq V$ образует в G клику, то в T существует узел $i \in I$, такой, что $A \subseteq X_i$, т. е. любая клика графа всегда целиком вложена в отдельный «мешок» дерева декомпозиции (\mathcal{X}, T) и, возможно, не в один;
- 2) если множества A и B индуцируют в G полный двудольный подграф, то в T имеется узел $i \in I$, такой, что $A \subseteq X_i$ или $B \subseteq X_i$, т. е. все вершины хотя бы одной доли этого подграфа принадлежат отдельному «мешку» дерева декомпозиции (\mathcal{X}, T) ;
- 3) пусть $w = \{i, j\}$ — произвольное ребро дерева T , а A и B — множества узлов двух компонент связности графа $T - w$, полученного удалением из дерева T ребра w . Тогда пересечение «мешков» $S = X_i \cap X_j$ образует (a, b) -сепаратор графа G при условии, что

$$a \in X_A = \bigcup_{l \in A} X_l \setminus S \neq \emptyset, \quad b \in X_B = \bigcup_{l \in B} X_l \setminus S \neq \emptyset.$$

Из данного утверждения следует, что если K_n — полный n -вершинный граф и K_{pq} — полный двудольный граф ($p + q = n$), то $\text{tw}(K_n) = n - 1$, $\text{tw}(K_{pq}) = \min(p, q)$. В общем случае при условии связности графа G верны естественные границы значений его древовидной ширины

$$0 < \text{tw}(G) \leq n - 1. \tag{1}$$

Очевидны следующие леммы.

Лемма 1. Пусть граф $G - v$ получен из графа $G = (V, E)$ удалением вершины $v \in V$. Тогда $\text{tw}(G - v) \leq \text{tw}(G)$.

Лемма 2. Пусть граф $G + e$ образован из графа $G = (V, E)$ добавлением ребра $e = \{a, b\}$, где $a, b \in V$. Тогда $\text{tw}(G) \leq \text{tw}(G + e)$.

Данные леммы свидетельствуют о том, что древовидная ширина графа — числовой параметр, которому свойственна монотонность относительно операций удаления вершины и добавления ребра: древовидная ширина графа не возрастает при удалении из него произвольной вершины и не убывает при добавлении в граф дополнительного ребра.

Следствие 1. Для любого подграфа $G' = G(A)$ графа $G = (V, E)$, индуцированного множеством $A \subseteq V$, верно $\text{tw}(G') \leq \text{tw}(G)$.

Следствие 2. Для каждой триангуляции $H = (V, E')$ графа $G = (V, E)$, где $E \subseteq E'$, всегда $\text{tw}(G) \leq \text{tw}(H)$.

Минором графа $G = (V, E)$ принято называть граф G' , который получается в результате применения к G одной или нескольких операций удаления вершины, удаления ребра или стягивания ребра. Ясно, что всякий индуцированный подграф графа G является его минором, однако обратное не всегда верно. Следующая лемма обобщает леммы 1 и 2.

Лемма 3 [1]. Для всякого минора G' графа G всегда $\text{tw}(G') \leq \text{tw}(G)$.

3. Графы с ограниченной древовидной шириной

Древовидная ширина отражает, насколько близок граф G к дереву, и, согласно утверждению 1, определяет размеры его клик и сепараторов: чем меньше $\text{tw}(G)$, тем ближе граф G к дереву и тем меньше у него по мощности клики и сепараторы. Так, все n -вершинные деревья ($n \geq 2$) имеют единичную древовидную ширину, размер всякой клики такого дерева равен 2, а каждый сепаратор — точка сочленения. Считается, что граф G обладает ограниченной древовидной шириной, если $\text{tw}(G) \leq k$ и k — положительная целая константа, не зависящая от n [9]. Например, если G — последовательно-параллельный граф, то $\text{tw}(G) \leq 2$, а для всякого графа Халина неизменно $\text{tw}(G) = 3$. Однако ограниченная древовидная ширина свойственна не всем графам. В частности, $\text{tw}(K_n) = n - 1$. Кроме того, существуют совсем простые по структуре графы, для которых значение $\text{tw}(G)$ растет с увеличением числа вершин, т. е. при $n \rightarrow \infty$ подобные графы все больше отдаляются от дерева. Типичным примером являются «сетки» — графы, построенные из правильных многогранников так, что любые два соседних многогранника имеют лишь одно общее ребро. Для «сетки» размера $n_1 \times n_2$ древовидная ширина вычисляется по формуле $\text{tw}(G) = \min(n_1, n_2) + 1$. При увеличении размера «сетки» ее древовидная ширина увеличивается, а число вершинной связности, наименьшая степень вершины, плотность и хроматическое число графа остаются неизменными.

4. Проблема вычисления древовидной ширины

Установить, имеет ли заданный граф ограниченную древовидную ширину, не всегда просто. В [3, 4] доказана NP-полнота следующей задачи: для графа G и целого положительного числа k верно ли, что $\text{tw}(G) \leq k$? Существует другая формулировка данной задачи, основанная на следующем утверждении.

Утверждение 2 [10]. Граф G имеет древовидную ширину не больше k тогда и только тогда, когда он является частичным k -деревом.

Использование понятия частичного k -дерева позволяет выявить структуру графа с ограниченной древовидной шириной. Определение частичного k -дерева базируется

на понятии k -дерева, которое, в свою очередь, определяется рекурсивно по правилам: полный граф из $k + 1$ вершин есть k -дерево; k -дерево с $i + 1$ вершинами получается из k -дерева с i вершинами путем добавления в него вершины v и k ребер таким образом, чтобы v стала смежной со всеми вершинами некоторой клики размера k (k -клики). Характерные особенности k -деревьев непосредственно вытекают из их построения и отражены в следующем утверждении.

Утверждение 3. Если граф $G = (V, E)$ есть k -дерево, то

- 1) G — связный граф;
- 2) G — хордальный граф и $\text{ch}(G) = m = |E|$;
- 3) G всегда имеет k -клику, но в нем нет $(k + 2)$ -клики;
- 4) $\varkappa(G) = \delta(G) = k$;
- 5) все максимальные клики в G есть $(k + 1)$ -клики;
- 6) $\varphi(G) = \chi(G) = k + 1$;
- 7) любой минимальный сепаратор в G является k -кликой;
- 8) $m = kn - k(k + 1)/2$.

Для того чтобы распознать k -дерево, нужно раскрутить рекурсию в обратном порядке. Это можно осуществить за полиномиальное время. Поскольку k -дерево — хордальный граф, то найти для него древовидную ширину не представляет труда (используя дерево клик).

Утверждение 4 [11]. Связный граф $G = (V, E)$ хордальный тогда и только тогда, когда для него существует дерево клик, определяемое следующим образом:

- 1) множество узлов дерева — множество $\{C_i : i \in I\}$ всех максимальных клик графа G ;
- 2) два узла C_i, C_j соединены ребром, если соответствующие им клики графа G имеют непустое пересечение, т. е. $C_i \cap C_j \neq \emptyset$;
- 3) для всякой вершины $v \in V$ графа G множество максимальных клик, содержащих эту вершину, индуцирует связный подграф, являющийся поддеревом дерева клик.

Известно, что любой хордальный граф имеет не более $n - 1$ максимальных клик [11]. Значит, дерево клик такого графа содержит не более $n - 1$ узлов и ребер. Хордальный граф может иметь несколько деревьев клик. Доказано [4], что каждое из них может быть найдено как остовное дерево наибольшего веса графа пересечений всех максимальных клик исходного графа. Здесь под весом ребра понимают число вершин, образующих пересечение подборающих максимальных клик. Таким образом, дерево клик хордального графа можно всегда построить за время $O(n^2)$, применяя известные алгоритмы построения оптимального остова [5, 8]. Очевидно, что для связного хордального графа G дерево клик с точностью до изоморфизма определяет оптимальное дерево декомпозиции, а мощность крупнейшей клики — древовидную ширину графа G : $\text{tw}(G) = \varphi(G) - 1$. По утверждению 3 плотность k -дерева равна $k + 1$ и потому $\text{tw}(G) = k$.

Следствие 3. Всякое k -дерево — хордальный граф с древовидной шириной, равной k .

Принадлежность k -деревьев к классу хордальных графов позволяет использовать свойства хордальных графов в определении древовидной ширины и построении деревьев декомпозиции графа.

Частичным k -деревом называют остовный подграф k -дерева [10]. Поскольку всякий n -вершинный граф является остовным подграфом графа K_n , то можно утверждать,

что подобный граф есть частичное $(n - 1)$ -дерево. Между тем в аспекте ограниченной древовидной ширины интерес представляет наименьшее значение k , при котором граф G — частичное k -дерево. Отыскать такое значение k — это значит найти наименьшее k -дерево, в которое можно вложить граф G . Таким образом, по следствию 3 для частичного k -дерева верно неравенство $\text{tw}(G) \leq k$, которое отвечает утверждению 2. Справедливость этого факта вытекает из монотонности древовидной ширины относительно операции добавления ребра в граф G , ведь всякое погружение графа G в граф H , удовлетворяющий требованиям k -дерева или хордальности, сводится к насыщению G дополнительными ребрами. Для триангуляций графа G справедливо

Утверждение 5 [10]. Если H является наименьшей триангуляцией графа G , то $\text{tw}(G) = \text{tw}(H)$. Для произвольной триангуляции H всегда $\text{tw}(G) \leq \text{tw}(H)$.

На сегодня сформировались два основных подхода к построению триангуляций графа: через удаление вершин, окрестность которых предварительно насыщается ребрами до клики, и через минимальные сепараторы, которые также дополняются до клики [12, 13]. В рамках этих подходов разработано много алгоритмов со временем работы $O(n^3)$.

Несмотря на то, что задача распознавания частичного k -дерева NP-полная [14], это никак не мешает вычислять древовидную ширину графа во многих ситуациях. Во-первых, известны точные неполиномиальные по времени алгоритмы нахождения $\text{tw}(G)$, основанные на методе динамического программирования и методе ветвей и границ [15, 16]. Разработаны также FPT-алгоритмы (*Fixed-Parameter Tractable algorithms*) [9, 17], способные при фиксированном k за время $O(2^k n^{O(1)})$ дать ответ на вопрос: $\text{tw}(G) \leq k$? Во-вторых, разработаны точные алгоритмы полиномиальной сложности для различных специальных классов графов [18–20]. В-третьих, выявлены необходимые полиномиально проверяемые признаки частичного k -дерева [10]. Так, если G — частичное k -дерево, то количество его ребер ограничено сверху: $m \leq kn - k(k + 1)/2$. В-четвертых, для $k = 1, 2$ найдены легко проверяемые критерии распознавания частичного k -дерева: $\text{tw}(G) \leq 1$ тогда и только тогда, когда граф G ациклический; $\text{tw}(G) \leq 2$ тогда и только тогда, когда каждый блок графа G является последовательно-параллельным графом [21]. В-пятых, к настоящему времени предложены различные схемы приближений [22, 23] и эвристические алгоритмы [24], позволяющие за полиномиальное время находить значения, близкие к истинному значению древовидной ширины графа. К сожалению, пока остается открытым вопрос о существовании полиномиальных схем приближения с гарантированной ошибкой. И наконец, актуальны границы возможных значений древовидной ширины графа и методы их уточнения.

5. Нижние оценки древовидной ширины и метод их уточнения

Можно назвать несколько причин, по которым полезны нижние оценки для древовидной ширины графа. Прежде всего, они могут быть использованы как критерии отсечения бесперспективных ветвей в методе ветвей и границ и в эвристических алгоритмах вычисления древовидной ширины графа. Кроме того, если в конкретной прикладной задаче для модельного графа G нижняя оценка $\text{tw}(G)$ слишком велика, то целесообразно отказаться от решения этой задачи методом динамического программирования на основе дерева декомпозиции. Иначе не исключена ситуация, когда время решения задачи может выйти за разумные пределы.

Рассмотрим нижние границы значений древовидной ширины графа G , связывающие $\text{tw}(G)$ с другими числовыми параметрами этого графа. Из свойств k -деревьев и

частичных k -деревьев (утверждений 2, 3 и следствия 3) непосредственно вытекают такие нижние оценки:

$$\delta(G) \leq \text{tw}(G); \tag{2}$$

$$\varkappa(G) \leq \text{tw}(G); \tag{3}$$

$$\varphi(G) - 1 \leq \text{tw}(G); \tag{4}$$

$$\chi(H) - 1 \leq \text{tw}(G); \tag{5}$$

$$\frac{2n - 1 - \sqrt{(2n - 1)^2 - 8m}}{2} \leq \text{tw}(G); \tag{6}$$

$$\frac{2n - 1 - \sqrt{(2n - 1)^2 - 8\text{ch}(G)}}{2} \leq \text{tw}(G). \tag{7}$$

Оценки (2)–(5) свидетельствуют о том, что если граф G имеет ограниченную древовидную ширину, т. е. $\text{tw}(G) \leq k$, то для него также ограничены сверху $\delta(G)$, $\varkappa(G)$, $\varphi(G)$, $\chi(H)$. Однако обратное, вообще говоря, неверно. Важно, что оценки (2), (3), (6) полиномиально вычисляемые. Но как раз они не очень хорошие.

Применение оценки (7) требует знания числа $\text{ch}(G)$, которое является результатом решения NP-трудной задачи наименьшего пополнения графа G до хордального. Для произвольной триангуляции H графа G всегда $\text{ch}(G) \leq \text{ch}(H)$. Поскольку всякая триангуляция H графа G по построению есть хордальный граф, то $\text{ch}(H)$ — число ребер в H . С учетом этого из (7) следует оценка

$$\frac{2n - 1 - \sqrt{(2n - 1)^2 - 8\text{ch}(H)}}{2} \leq \text{tw}(G). \tag{8}$$

Конечно, чем меньше использовано ребер для пополнения графа G до H , тем выше нижняя граница значений $\text{tw}(G)$ в (8). Поэтому при вычислении оценки (8) следует использовать триангуляции хорошего качества, для построения которых могут быть применены различные известные полиномиальные по времени приближенные и эвристические алгоритмы [22, 24].

Оценки (4) и (5), выраженные через плотность и хроматическое число графа, трудновычисляемые. На первый взгляд может показаться, что именно плотность и хроматическое число тесно связаны с древовидной шириной графа, и если значения $\varphi(G)$ и $\chi(G)$ невелики, то невелико и значение $\text{tw}(G)$. Однако на самом деле это не так: разности $\text{tw}(G) - \varphi(G)$, $\text{tw}(G) - \chi(G)$ могут быть величинами, зависящими от n . Например, «сетка» из треугольников размера $n_1 \times n_2$ неизменно имеет плотность и хроматическое число, равные 3, при этом значение древовидной ширины зависит от размера «сетки». Существуют графы без треугольников с произвольно большим хроматическим числом [5, с. 242]. Правила построения последовательности $G_2, G_3, \dots, G_i, \dots$ таких графов указал А. А. Зыков в 1949 г. В этой последовательности $G_2 = K_2$, а каждый граф G_i , $i = 3, 4, \dots$, имеет $n_i = 3 \cdot 2^{i-2} - 1$ вершин, $\varphi(G_i) = 2$, $\chi(G_i) = i$, $\text{tw}(G_i) = (n_i - 1)/2$. Поэтому значения разностей $\text{tw}(G_i) - \varphi(G_i)$, $\text{tw}(G_i) - \chi(G_i)$ увеличиваются с ростом i .

Полиномиально вычисляемые оценки (2), (3), связанные со степенями вершин графа, можно усилить. Для этого следует воспользоваться монотонностью $\text{tw}(G)$ и немонотонностью $\delta(G)$ и $\varkappa(G)$ относительно операции удаления вершины.

Лемма 4. Пусть $\Psi(G)$ — множество всех индуцированных подграфов графа G . Тогда

$$\max_{G' \in \Psi(G)} \delta(G') \leq \text{tw}(G); \quad (9)$$

$$\max_{G' \in \Psi(G)} \varkappa(G') \leq \text{tw}(G). \quad (10)$$

Доказательство. В самом деле, согласно следствию 1 и оценке (2), для каждого индуцированного подграфа $G' \in \Psi(G)$ графа G справедливы неравенства

$$\delta(G') \leq \text{tw}(G') \leq \text{tw}(G).$$

Отсюда вытекают соотношения

$$\max_{G' \in \Psi(G)} \delta(G') \leq \max_{G' \in \Psi(G)} \text{tw}(G') \leq \text{tw}(G)$$

и правильность (9). Аналогичным образом с использованием (3) доказывается (10). ■

В общем случае множество $\Psi(G)$ содержит порядка 2^n индуцированных подграфов графа G , и их перебор может потребовать слишком много времени. Поэтому при больших значениях n в (9) и (10) вместо множества $\Psi(G)$ целесообразно использовать некоторое его подмножество, которое имеет небольшую мощность и отражает структурные особенности графа G через его части. Таким подмножеством, например, может быть множество $\mathfrak{P}(G) \subseteq \Psi(G)$, называемое колодой графа G . Колода n -вершинного графа $G = (V, E)$ состоит из n его подграфов $G - v$, полученных из G удалением только одной вершины $v \in V$. Для $\mathfrak{P}(G)$ оценки (9) и (10) полиномиально вычислимы. При формировании колоды $\mathfrak{P}(G)$ удалению подлежит каждая вершина графа G , включая специальные вершины, такие, как симплициальные вершины и точки сочленения, если они имеются в G . Примечательно, что чем больше специальных вершин в графе G , тем больше шансов улучшить нижние оценки (2), (3) на основе $\mathfrak{P}(G)$. Наибольший эффект колоды обеспечивают симплициальные вершины.

Лемма 5. Если $v \in V$ есть симплициальная вершина графа $G = (V, E)$ степени $\deg(v)$, то

$$\max(\deg(v), \text{tw}(G - v)) = \text{tw}(G). \quad (11)$$

Справедливость данной леммы следует из утверждения 7, которое приведено и доказано ниже. Равенство (11) дает для симплициальной вершины v соотношения

$$\max(\deg(v), \delta(G - v)) \leq \text{tw}(G), \quad \max(\deg(v), \varkappa(G - v)) \leq \text{tw}(G).$$

Таким образом, при вычислении оценок (9), (10) по $\mathfrak{P}(G)$ для каждой симплициальной вершины v вместо $\delta(G - v)$ и $\varkappa(G - v)$ необходимо применять соответствующие величины

$$\max(\deg(v), \delta(G - v)), \quad \max(\deg(v), \varkappa(G - v)).$$

Распознается симплициальная вершина за время $O(n^2)$.

Для уточнения оценок (2) и (3) в качестве представительного подмножества множества $\Psi(G)$ можно рассматривать также множество $\mathfrak{B}(G)$ блоков графа G . Это множество получается разделением графа G на части с помощью точек сочленения. Напомним, что блок графа G — максимальный относительно включения связный его подграф

без точек сочленения. Здесь важно отметить, что каждый блок является индуцированным подграфом графа G (значит, к нему применимо следствие 1), $|\mathfrak{B}(G)| \leq n$ и всякий блок можно отыскать в G за полиномиальное время. Конечно, если граф G состоит только из одного блока (таким, например, является «сетка» из треугольников), то предложенный выше алгоритмический метод улучшения нижних оценок (2), (3) не дает эффекта. В подобных случаях следует искать другие правила учета специальных вершин графа и выделения в $\Psi(G)$ представительного подмножества или оценивать значения древовидной ширины сверху.

6. Верхнее конструктивное оценивание и предобработка графа

Явных верхних оценок сравнительно немного. К ним можно отнести тривиальную оценку (1). Известны также верхние оценки для некоторых специальных классов графов. Например, для всякого планарного графа G с диаметром $d(G)$ выполняется соотношение [25] $tw(G) \leq 3d(G) - 2$. В общем случае на практике для определения верхних границ древовидной ширины графа чаще всего используют конструктивное оценивание на основе неравенства

$$tw(G) \leq tw(H), \quad (12)$$

которое, согласно следствию 2 и утверждению 5, верно для любой триангуляции H графа G . Здесь, как и в оценке (8), нужны качественные триангуляции.

Тривиальную оценку (1) и верхние оценки, найденные с помощью триангуляций, можно уточнить с помощью процедуры предобработки, основанной на принципе «разделяй и властвуй». Цель предобработки — разделить исходный граф G на части таким образом, чтобы задача нахождения наименьшей триангуляции H для графа G свелась к решению той же самой задачи, но только для графов меньшей размерности. Затем в зависимости от размера полученных частей графа использовать для построения качественных триангуляций точные, приближенные или эвристические алгоритмы. Один из возможных подходов надлежащего разбиения графа — выделение частей графа с помощью сепараторов.

Пусть $\mathfrak{S}(G) = \{S_j : j \in I_s\}$ — множество сепараторов графа $G = (V, E)$. Предположим сначала, что всякий сепаратор состоит лишь из одной вершины — точки сочленения графа G . Рассмотрим семейство $\mathfrak{B}(G) = \{B_i : i \in I_b\}$ блоков, полученных путем разделения графа точками сочленения. Пусть $\{X_i : i \in I_b\}$ — семейство блоковых множеств, где $X_i \subseteq V$ состоит из вершин блока B_i , и только из них ($i \in I_b$). Будем опираться на известные свойства блоков [5]:

- 1) любые два блока имеют не более одной общей вершины, а каждое ребро графа входит только в один его блок;
- 2) если блок графа содержит вершины a и b , то он также содержит и всякую простую (a, b) -цепь этого графа;
- 3) семейство блоковых множеств является покрытием множества вершин графа. Каждая пара блоковых множеств либо не пересекается, либо имеет единственную общую вершину, и эта вершина — точка сочленения графа.

Соотнесем графу G двудольный граф $T = (I, W)$ с множеством узлов $I = I_b \cup I_s$, в котором два узла i и j смежны, если узел i соответствует блоку B_i , а узел j — сепаратору S_j и $S_j \subset X_i$ ($i \in I_b, j \in I_s$). Такой граф описывает структуру графа G с точки зрения взаимного расположения его блоков и точек сочленения. Исходя из перечисленных выше свойств блоков, нетрудно убедиться, что если граф G связан, то T —

всегда дерево [5]. Это дерево принято называть деревом блоков и точек сочленения графа G . На его основе построим дерево декомпозиции графа G . Для этого сопоставим каждому узлу дерева T «мешок» по следующим правилам: пусть для узла $i \in I_b$ в роли «мешка» выступает блоковое множество X_i , а для узла $j \in I_s$ — множество вершин сепаратора S_j . Таким образом, «мешки» для узлов-сепараторов будут содержать только одну вершину — соответствующую точку сочленения графа G .

Утверждение 6. Пара (\mathcal{X}, T) , где $\mathcal{X} = \{X_i : i \in I_b\} \cup \{S_j : j \in I_s\}$ и $T = (I, W)$ — дерево блоков и точек сочленения графа G , определяет дерево декомпозиции этого графа.

Доказательство. Достаточно проверить выполнимость трех требований, которым должно удовлетворять всякое дерево декомпозиции. Доказательство опирается на свойства блоков. В самом деле, в качестве «мешков» здесь выступают блоковые множества и одновершинные сепараторы. Каждый из таких сепараторов принадлежит двум и более блоковым множествам. Согласно свойству 3, семейство блоковых множеств покрывает все вершины графа G . Исходя из свойства 1, всякое ребро $\{a, b\}$ графа G входит в некоторый блок, а значит, оба конца данного ребра принадлежат «мешку», отвечающему этому блоку. На основании свойств 1 и 3 вершина $v \in V$ графа G , если она не является точкой сочленения этого графа, всегда принадлежит лишь одному блоковому множеству (одному «мешку») и, следовательно, образует поддереву дерева T , состоящее из одного узла. Если вершина v — точка сочленения графа G , то по построению дерева блоков и точек сочленения в T всегда существует узел $j \in I_s$, смежный с узлами $i_1, i_2, \dots, i_\tau \in I_b$, которые соответствуют некоторому подмножеству блоков $B' \subseteq \mathfrak{B}(G)$ графа G , причем точка сочленения v принадлежит «мешкам» только этих блоков и никаким другим. Таким образом, все узлы дерева, «мешки» которых содержат точку сочленения v , индуцируют связный подграф, являющийся поддеревом дерева T . Все три требования, предъявляемые к дереву декомпозиции, выполнены. ■

Заметим, что дерево декомпозиции, определенное через блоковые множества и точки сочленения графа G , не имеет кратных «мешков», но обязательно имеет вложенные «мешки». Непустые пересечения «мешков» этого дерева задают все одновершинные сепараторы графа G . Ширина данного дерева устанавливает верхнюю границу значений $\text{tw}(G)$. К сожалению, если граф G является 2-связным, то эта граница не отличается от границы, заданной неравенством (1). Если граф G содержит точки сочленения, т. е. является 1-связным, то для $\mathfrak{B}(G)$ верна формула

$$\text{tw}(G) = \max\{\text{tw}(B_i) : i \in I_b\}, \quad (13)$$

которая означает, что древовидная ширина графа G может быть найдена через значения древовидной ширины всех его блоков. Пусть для каждого блока B_i определена некоторая его триангуляция H_i ($i \in I_b$). Тогда очевидна верхняя оценка

$$\text{tw}(G) \leq \max\{\text{tw}(H_i) : i \in I_b\}. \quad (14)$$

Поскольку при разложении графа на блоки размерность задачи построения наименьшей триангуляции снижается, то область действия погрешностей, вносимых схемами приближений и эвристиками, сокращается. Возможно также применение точных методов для блоков с небольшим числом вершин. Улучшение качества триангуляций H_i позволяет в силу (14) совершенствовать оценку (12).

Деревья декомпозиции, построенные для блоков, можно «склеить» через соответствующие сепараторы в графовую структуру, которая по утверждению 6 является

деревом декомпозиции исходного графа G . Заметим, что в данном случае все сепараторы, образующие $\mathfrak{S}(G)$, являются минимальными и попарно непересекающимися, так как они одновершинные. Семейство $\mathfrak{B}(G) = \{B_i : i \in I_b\}$ можно создавать поэтапно — путем нахождения очередной точки сочленения в какой-либо компоненте связности графа, полученной на предыдущем этапе разбиения. При этом состав блоков в $\mathfrak{B}(G)$ не зависит от порядка выбора точек сочленения.

Рассмотренный подход к предобработке графа можно обобщить на произвольные сепараторы. Если граф $G = (V, E)$ является l -связным, то $\chi(G) \geq l$ и в G существуют сепараторы размера l . Пусть $\mathfrak{S}(G)$ — множество сепараторов графа G и $S \in \mathfrak{S}(G)$ — некоторый сепаратор размера $l = |S|$. Обозначим через Y_1, Y_2, \dots, Y_τ области связности графа $G(V \setminus S)$. Части G_1, G_2, \dots, G_τ графа G , как результат разбиения его сепаратором S , определим сначала как подграфы, индуцированные множествами вершин $X_i = Y_i \cup S$ ($i \in I = \{1, 2, \dots, \tau\}$). Для сохранения l -связности к каждой части добавим всевозможные ребра между вершинами множества S . Таким образом, если через $K(S)$ обозначить полный граф на множестве вершин S , то окончательно $G_i = G(Y_i \cup S) + K(S)$, т. е. в каждом подграфе G_i вершины сепаратора S образуют клику ($i \in I$). Сепаратор S графа G назовем безопасным сепаратором относительно $\text{tw}(G)$, если

$$\text{tw}(G) = \max\{\text{tw}(G_i) : i \in I\}. \quad (15)$$

Лемма 6. Для произвольного сепаратора S графа G всегда

$$\text{tw}(G) \leq \max\{\text{tw}(G_i) : i \in I\}. \quad (16)$$

Доказательство. Предположим, что значение максимума в правой части (16) равно μ . Тогда для каждого подграфа G_i можно построить дерево декомпозиции ширины не более μ ($i \in I$). По высказыванию 1 утверждения 1 в этом дереве обязательно найдется узел, «мешок» которого содержит сепаратор S , так как вершины этого сепаратора образуют в G_i клику. Дерево декомпозиции графа G может быть сформировано путем добавления в каждое дерево декомпозиции, построенное для G_i ($i \in I$), дополнительного узла с «мешком», равным S , и «склеиванием» данных деревьев посредством этого дополнительного узла. Очевидно, что ширина результирующего дерева декомпозиции не будет превышать μ . ■

Утверждение 7. Если сепаратор S графа G — клика, то он безопасный относительно $\text{tw}(G)$.

Доказательство. Если $G(S)$ — клика, то для любого подграфа G_i , как части разбиения G сепаратором S , справедливо равенство

$$G_i = G(Y_i \cup S) + K(S) = G(Y_i \cup S), \quad i \in I,$$

которое означает, что G_i — индуцированный подграф графа G . Исходя из следствия 1, $\text{tw}(G_i) \leq \text{tw}(G)$. Отсюда

$$\max\{\text{tw}(G_i) : i \in I\} \leq \text{tw}(G). \quad (17)$$

Из неравенств (16), (17) вытекает равенство (15). ■

Из утверждения 7 следует справедливость леммы 5, так как окрестность всякой симплициальной вершины — сепаратор, являющийся кликой и отделяющий эту вершину от других вершин графа. Утверждение 7 определяет достаточные условия безопасности сепаратора относительно древовидной ширины графа. Одновершинные сепараторы отвечают данному условию как одновершинные полные графы. Это является подтверждением равенства (13).

Следствие 4. Все одновершинные сепараторы графа G являются безопасными относительно $\text{tw}(G)$.

Условие утверждения 7 можно расширить. Сепаратор S графа G есть почти клика, если существует вершина $v \in S$, такая, что множество $S - v$ образует клику графа G .

Лемма 7 [3]. Сепаратор S графа G является минимальным тогда и только тогда, когда для каждой вершины $v \in S$ и всякой области связности Y_i ($i \in I$) графа $G(V \setminus S)$ всегда найдется вершина $w \in Y_i$, смежная с v в $G(Y_i \cup S)$.

Утверждение 8. Если S — почти клика и минимальный сепаратор графа G , то он безопасный относительно $\text{tw}(G)$.

Доказательство. Пусть вершина $v \in S$ такова, что множество $S - v$ образует клику графа G . Покажем, что каждая компонента связности $G(Y_i \cup S)$ графа $G(V \setminus S)$ содержит $K(S)$ в качестве минора ($i \in I$). Рассмотрим вершину v минимального сепаратора S , которая не входит в клику этого сепаратора. Согласно лемме 7, существует вершина $w \in Y_i$, смежная с v в $G(Y_i \cup S)$. Для всякой другой вершины из $S - v$ также имеется смежная вершина в $G(Y_i \cup S)$. Исходя из этого, все вершины области связности Y_i можно стянуть к v . В результате стягивания получим граф $K(S)$. Убедимся теперь в справедливости равенства (15). Кроме $G(Y_i \cup S)$, проанализируем еще одну компоненту связности графа $G(V \setminus S)$. Пусть это будет компонента $G(Y'_i \cup S)$. Она так же, как и $G(Y_i \cup S)$, включает $K(S)$ в качестве минора. Между тем граф $G - Y_i$ содержит $G(Y'_i \cup S)$ в качестве подграфа, а значит, $K(S)$ в качестве минора. Отсюда следует, что граф $G_i = G(Y_i \cup S) + K(S)$ является минором для G . Поэтому по лемме 3 верно неравенство (17) и, учитывая лемму 6, справедливо равенство (15). ■

Следствие 5. Все минимальные двухвершинные сепараторы графа G безопасные относительно $\text{tw}(G)$.

Справедливость следствия 5 непосредственно вытекает из утверждения 8, так как всякое двухэлементное множество вершин графа G образует почти клику.

Результаты утверждений 6–8 и следствий 4, 5 могут быть использованы для уточнения верхних оценок древовидной ширины исходного графа G следующим образом:

- 1) первоначально семейство $\mathfrak{B}(G)$ состоит только из G ;
- 2) для каждого элемента из $\mathfrak{B}(G)$ — некоторой части исходного графа — находим безопасный сепаратор. Поиск начинаем с сепараторов размера 1, затем, если таких нет, выполняем поиск минимальных сепараторов размера 2 и т. д.;
- 3) если безопасный сепаратор найден, то рассматриваемую часть подвергаем разбиению и полученные более мелкие части помещаем в $\mathfrak{B}(G)$. Если безопасный сепаратор не найден, то этот элемент из $\mathfrak{B}(G)$ далее уже не подвергается обработке;
- 4) шаги 2 и 3 повторяем до тех пор, пока в $\mathfrak{B}(G)$ не окажется частей с безопасными сепараторами;
- 5) далее применяем алгоритмы триангуляции к элементам из $\mathfrak{B}(G)$, если это необходимо, и затем соотношение (14).

Для того чтобы описанный процесс можно было реализовать за полиномиальное время, необходимо на шаге 2 осуществлять поиск только тех сепараторов, которые легко вычисляются. К ним, например, относятся сепараторы размера 1 и 2. Они вычисляются с помощью поиска в глубину [5] за время $O(n + m)$. Для нахождения сепараторов размера 3 и более можно привлекать потоковые алгоритмы (Форда — Фалкерсона, Эд-

монса — Карпа и др.) [8]. Однако эти алгоритмы более затратные по времени, хотя и являются полиномиальными.

Следует отметить, что результат разбиения исходного графа на части безопасными сепараторами, вообще говоря, зависит от порядка выбора этих сепараторов, поскольку при $l > 1$ в l -связном графе допустимы зависимые (разделяющие друг друга) сепараторы мощности l , и после разбиения по одному из них невозможно уже произвести разбиение по второму [26]. Тем не менее если использовать только безопасные сепараторы, то на каждом шаге разбиения верно (15) для любого построенного варианта $\mathfrak{B}(G)$. Многовариантность может сказаться только на построении триангуляций для элементов из $\mathfrak{B}(G)$ и, следовательно, на значении верхней границы древовидной ширины графа G в (14). Конечно, существуют графы, в которых нет минимальных сепараторов, являющихся кликами или почти кликами. Для них предложенный выше алгоритмический метод улучшения верхней оценки древовидной ширины не дает эффекта. В таких случаях требуются другие безопасные сепараторы и иные методы предварительной обработки графа.

Заключение

В настоящее время проблеме вычисления древовидной ширины графа и построения дерева декомпозиции в зарубежных научных изданиях уделяется много внимания. Использование дерева декомпозиции при решении NP-трудных графовых задач — один из современных подходов преодоления высокой вычислительной сложности этих задач. В российских изданиях тематика древовидной ширины графов освещена весьма скромно. Имеются работы по исследованию структуры сепараторов в многосвязных графах [26, 27]. Частично введена соответствующая терминология в словарь по теории графов [28]. Изучаются хордальные графы [29] и ведется разработка новых алгоритмов построения минимальных триангуляций [30]. Имеются работы по применению древовидной ширины графа в решении задач дискретной оптимизации [31]. Однако эти работы не охватывают весь спектр открытых вопросов, связанных с древовидной шириной графа.

В данной работе дан краткий обзор современных результатов по проблеме вычисления древовидной ширины. Исследованы нижние и верхние оценки, связывающие древовидную ширину с другими числовыми параметрами графа, проанализировано их качество и сложность вычисления. Предложены и теоретически обоснованы алгоритмические методы улучшения этих оценок, основанные на немонотонности отдельных числовых параметров графа относительно операции удаления вершин графа и разложении графа сепараторами. Данные методы указывают направления исследований, перспективные для развития алгоритмических и прикладных аспектов теории графов.

ЛИТЕРАТУРА

1. Robertson N. and Seymour P. D. Graph minors. II. Algorithmic aspects of treewidth // J. Algorithms. 1986. V. 7. P. 309–322.
2. Kleinberg J. and Tardos E. Algorithm Design. Boston: Addison-Wesley, 2005.
3. Bodlaender H. L. and Thilikos D. M. Treewidth for graphs with small chordality // Disc. Appl. Math. 1997. V. 79. P. 45–61.
4. Parra A. and Scheffler P. Characterizations and algorithmic applications of chordal graph embeddings // Disc. Appl. Math. 1997. V. 79. P. 171–188.
5. Емеличев В. А., Мельников О. И., Сарванов В. И., Тышкевич Р. И. Лекции по теории графов. М.: Наука, 1990.

6. *Rose D., Tarjan R. E., and Lueker G.* Algorithmic aspects of vertex elimination on graphs // SIAM J. Comput. 1976. V. 5. P. 146–160.
7. *Buneman P.* A characterization of rigid circuit graphs // Disc. Math. 1974. V. 9. P. 205–212.
8. *Быкова В. В.* Дискретная математика с использованием ЭВМ. Красноярск: РИО КрасГУ, 2006.
9. *Bodlaender H. L.* Some classes of graphs with bounded treewidth // Technical Report RUU-CS-76-22, Dept. Comput. Science, Univ. Utrecht, 1998.
10. *Bodlaender H. L.* A partial k -arboretum of graphs with bounded treewidth // Theor. Comput. Sci. 1998. V. 209. P. 1–45.
11. *Blair J. R. S. and Peyton B.* An introduction to chordal graphs and clique trees // Graph theory and sparse matrix computation. New York: Springer, 1993. P. 1–29.
12. *Berry A.* A wide-range efficient algorithm for minimal triangulation // Proc. Tenth Annual ACM-SIAM Symposium on Disc. Algorithms. Philadelphia: SIAM, 1999. P. 860–861.
13. *Berry A., Heggernes P., and Simonet G.* The minimum degree heuristic and the minimal triangulation process // LNCS. 2003. V. 2880. P. 58–70.
14. *Arnborg S., Corneil D. G., and Proskurowski A.* Complexity of finding embeddings in a k -tree // SIAM J. Alg. Disc. Meth. 1987. V. 8. P. 277–284.
15. *Gogate V. and Dechter R.* A complete anytime algorithm for treewidth // Proc. 20Th Conference on Uncertainty in Artificial Intelligence. Arlington, Virginia, USA: AUAI Press, 2004. P. 201–208.
16. *Bodlaender H. L., Fomin F. V., Koster A. M. C. A., et al.* On exact algorithms for treewidth // LNCS. 2006. V. 4168. P. 672–683.
17. *Bodlaender H. L. and Kloks T.* Efficient and constructive algorithms for the pathwidth and treewidth of graphs // J. Algorithms. 1996. V. 21. P. 358–402.
18. *Bodlaender H. L. and Rotics U.* Computing the treewidth and the minimum fill-in with the modular decomposition // Algorithmica. 2003. V. 36. P. 375–408.
19. *Broersma H., Dahlhaus E., and Kloks T.* A linear time algorithm for minimum fill-in and tree width for distance hereditary graphs // Disc. Appl. Math. 2000. V. 99. P. 367–400.
20. *Broersma H., Kloks T., Kratsch D., and Muller H.* A generalization of AT-free graphs and a generic algorithm for solving triangulation problems // Algorithmica. 2002. V. 32. P. 594–610.
21. *Bodlaender H. L. and Fluiter B.* Parallel algorithms for series parallel graphs and graphs with treewidth two // Algorithmica. 2001. V. 29. P. 543–559.
22. *Amir E.* Efficient approximations for triangulation of minimum treewidth // Proc. 17Th Conference on Uncertainty in Artificial Intelligence. CA, USA: Morgan Kaufmann Publishers Inc. San Francisco, 2001. P. 7–15.
23. *Bouchitte V., Kratsch D., Muller H., and Todinca I.* On treewidth approximations // Disc. Appl. Math. 2004. V. 6. P. 183–196.
24. *Clautiaux F., Moukrim A., Negre S., and Carlier J.* Heuristic and meta-heuristic methods for computing graph treewidth // RAIRO Oper. Res. 2004. V. 38. P. 13–26.
25. *Eppstein D.* Diameter and treewidth in minor-closed families // Algorithmica. 2000. V. 27. P. 275–291.
26. *Карпов Д. В.* Разделяющие множества в k -связном графе // Зап. научн. семин. ПОМИ. 2006. Т. 340. С. 33–60.
27. *Карпов Д. В., Пастор А. В.* О структуре k -связного графа // Зап. научн. семин. ПОМИ. 2000. Т. 266. С. 76–106.
28. *Евстигнеев А. А., Касьянов В. Н.* Словарь по графам в информатике. Новосибирск: ООО «Сибир. научн. изд-во», 2000.

-
29. *Турсунбай кызы Б.* Деревья клик хордального графа и деревья подграфов // Конструирование и оптимизация программ. Новосибирск: Институт систем информатики СО РАН, 2008. Вып. 16. С. 314–321.
 30. *Быкова В. В., Никульская Н. А.* Алгоритмические аспекты минимальных триангуляций графа // Труды XIV Межд. конф. по эвент. матем. и смежным вопросам. Красноярск: КГТЭИ, СФУ, 2010. С. 26–32.
 31. *Щербина О. А.* Локальные элиминационные алгоритмы решения разреженных дискретных задач // Журн. вычисл. матем. и матем. физ. 2008. Т. 48. № 1. С. 159–175.

**ВЫЧИСЛИТЕЛЬНАЯ СЛОЖНОСТЬ ЗАДАЧИ
АППРОКСИМАЦИИ ГРАФАМИ С КОМПОНЕНТАМИ СВЯЗНОСТИ
ОГРАНИЧЕННОГО РАЗМЕРА**

В. П. Ильев, А. А. Навроцкая

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

E-mail: iljev@mail.ru, nawrocki@ya.ru

Рассматриваются новые варианты задачи аппроксимации графа, в которых имеются ограничения на размер компонент связности аппроксимирующих графов. Доказано, что если в качестве последних допускаются графы с компонентами связности мощностей $1, 2, \dots, p \geq 3$, то задача аппроксимации графа является *NP*-трудной, а в случае $p = 2$ она полиномиально разрешима.

Ключевые слова: *аппроксимация графа, полиномиально разрешимая задача, NP-трудная задача.*

Введение

Задачи аппроксимации графов возникают при анализе систем взаимосвязанных объектов, в частности в задачах классификации. При этом минимизируется число связей между классами и число недостающих связей внутри классов. В литературе рассматривались задачи, в которых количество классов не ограничено, ограничено сверху или равно наперед заданному числу. Постановки и различные интерпретации этих задач можно найти в [1–5]. В настоящей работе рассматриваются варианты задачи, в которых имеются ограничения на мощности классов.

Граф без петель и кратных ребер называется *M-графом*, если каждая его компонента связности есть полный граф. Обозначим через $\mathcal{M}(V)$ множество всех *M-графов* на множестве вершин V , а через $\mathcal{M}^p(V)$ — множество всех *M-графов* на множестве V , в которых мощность каждой компоненты связности равна p , $2 \leq p \leq |V|$. Будем говорить, что *M-граф* принадлежит множеству $\mathcal{M}^{\leq p}(V)$, если мощность каждой его компоненты не превышает целого числа p , $2 \leq p \leq |V|$.

Пусть $G_1 = (V, E_1)$ и $G_2 = (V, E_2)$ — помеченные графы, тогда *расстояние* между ними определяется следующим образом: $\rho(G_1, G_2) = |E_1 \Delta E_2|$, где $E_1 \Delta E_2 = (E_1 \setminus E_2) \cup (E_2 \setminus E_1)$, т. е. $\rho(G_1, G_2)$ равно числу несовпадающих ребер в графах G_1, G_2 .

ЗАДАЧА АППРОКСИМАЦИИ ГРАФА (A). Дан произвольный n -вершинный граф $G = (V, E)$. Найти такой *M-граф* $M^* \in \mathcal{M}(V)$, что $\rho(G, M^*) = \min_{M \in \mathcal{M}(V)} \rho(G, M)$.

ЗАДАЧА $A^{\leq p}$. Дан n -вершинный граф $G = (V, E)$ и целое число p , $2 \leq p \leq n$. Найти такой граф $M^* \in \mathcal{M}^{\leq p}(V)$, что $\rho(G, M^*) = \min_{M \in \mathcal{M}^{\leq p}(V)} \rho(G, M)$.

ЗАДАЧА A^p . Дан граф $G = (V, E)$, такой, что $|V| = pq$, где p, q — целые положительные числа. Найти такой граф $M^* \in \mathcal{M}^p(V)$, что $\rho(G, M^*) = \min_{M \in \mathcal{M}^p(V)} \rho(G, M)$.

В работе [6] показано, что задача **A** *NP*-трудна. Задачи $A^{\leq p}$ и A^p ранее не рассматривались. В настоящей работе доказано, что задачи $A^{\leq 2}$ и A^2 полиномиально разрешимы, а для любого фиксированного $p \geq 3$ задачи $A^{\leq p}$ и A^p являются *NP*-трудными.

1. Полиномиально разрешимые случаи

Лемма 1. Для любого графа $G = (V, E)$ существует такое оптимальное решение $M^* = (V, E^*)$ задачи $\mathbf{A}^{\leq p}$ при $2 \leq p \leq 3$, что $E^* \subseteq E$.

Доказательство. При $p = 2$ утверждение леммы верно для любого из оптимальных M -графов, т. е. для любого M -графа, являющегося оптимальным решением задачи $\mathbf{A}^{\leq 2}$. Действительно, допустим, что существует оптимальный M -граф $M' = (V, E') \in \mathcal{M}^{\leq 2}(V)$, для которого утверждение леммы неверно. Значит, в нем найдется компонента связности, состоящая из двух вершин u и v , таких, что $uv \notin E$. Тогда рассмотрим M -граф $M^* = (V, E' \setminus \{uv\})$. Получаем, что $\rho(G, M^*) = \rho(G, M') - 1$, что противоречит оптимальности M -графа M' .

Докажем утверждение для $p = 3$. Пусть $M' = (V, E')$ — оптимальное решение задачи $\mathbf{A}^{\leq 3}$, не удовлетворяющее утверждению леммы. Значит, $E' \setminus E \neq \emptyset$. Допустим, в M' существует компонента связности, содержащая более одного ребра из множества $E' \setminus E$, т. е. существуют такие вершины u, v, w , что $uv, vw \in E' \setminus E$. Рассмотрим M -граф $M = (V, E' \setminus \{uv, vw\})$, тогда $\rho(G, M) = \rho(G, M') - 2$, что противоречит оптимальности M -графа M' .

Аналогично доказывается, что в графе M' нет компонент связности, содержащих ребро из $E' \setminus E$ и состоящих из двух вершин.

Таким образом, в графе M' существует ровно $|E' \setminus E|$ компонент связности, в каждой из которых есть три вершины u, v, w , таких, что $uv, vw \in E$, $uw \in E' \setminus E$. Тогда рассмотрим граф $M^* = (V, E^*)$, в котором все такие компоненты разделены на две: смежные вершины u, v составляют отдельную компоненту связности, а вершина w — другую. Так как удалены все ребра множества $E' \setminus E$, то $E^* \subseteq E$. Очевидно, что $\rho(G, M^*) = \rho(G, M')$. Следовательно, граф M^* также является оптимальным и удовлетворяет условию леммы. ■

Теорема 1. Задача $\mathbf{A}^{\leq 2}$ полиномиально разрешима.

Доказательство. Докажем, что задача $\mathbf{A}^{\leq 2}$ эквивалентна задаче о наибольшем паросочетании. Пусть дан граф $G = (V, E)$. Из леммы 1 следует, что задача $\mathbf{A}^{\leq 2}$ сводится к отысканию M -графа $M^* = (V, E^*) \in \mathcal{M}^{\leq 2}(V)$, являющегося подграфом графа $G = (V, E)$, в котором множество ребер E^* имеет максимальную мощность. Так как $M^* \in \mathcal{M}^{\leq 2}(V)$, то никакие два ребра из $E^* \subseteq E$ не имеют общей вершины; другими словами, E^* — паросочетание графа G . Таким образом, найдя наибольшее паросочетание в графе G , мы получим оптимальное решение $M^* = (V, E^*)$ задачи $\mathbf{A}^{\leq 2}$ на графе G , и наоборот. Хорошо известно, что задача поиска наибольшего паросочетания в произвольном графе полиномиально разрешима [7]. ■

Следствие 1. Задача $\mathbf{A}^{\leq 3}$ на графах, не содержащих полных трехвершинных подграфов, полиномиально разрешима.

Доказательство. Пусть дан граф $G = (V, E)$, не содержащий полных трехвершинных подграфов. Из леммы 1 следует, что существует такое оптимальное решение $M^* = (V, E^*)$ задачи $\mathbf{A}^{\leq 3}$ на графе G , что M^* является подграфом графа G , а так как G не содержит треугольников, то $M^* \in \mathcal{M}^{\leq 2}(V)$. Значит, задача $\mathbf{A}^{\leq 3}$ на графах, не содержащих полных трехвершинных подграфов, эквивалентна задаче $\mathbf{A}^{\leq 2}$. ■

Теорема 2. Задача \mathbf{A}^2 полиномиально разрешима.

Доказательство. Фиксируем такой граф $G = (V, E)$, что $|V| = 2q$. Рассмотрим M -граф $M^* = (V, E^*)$ — произвольное допустимое решение задачи \mathbf{A}^2 . Из того, что

$M^* \in \mathcal{M}^2(V)$, следует $|E^*| = q$. Вычислим расстояние между графами G и M^* :

$$\rho(G, M^*) = |E \Delta E^*| = |E| + |E^*| - 2|E \cap E^*| = |E| + q - 2|E \cap E^*|.$$

Поскольку $|E|$ и q фиксированы, то $\rho(G, M^*)$ тем меньше, чем $|E \cap E^*|$ больше. Очевидно, что $E \cap E^*$ — паросочетание в графе G . Таким образом, задача \mathbf{A}^2 сводится к нахождению наибольшего паросочетания в графе G . ■

2. NP -трудные задачи

Докажем, что задачи $\mathbf{A}^{\leq p}$ и \mathbf{A}^p являются NP -трудными для любого фиксированного $p \geq 3$.

Лемма 2. Пусть $M = (V, E) \in \mathcal{M}^{\leq p}(V)$, где $|V| = pq$. Тогда

$$|E| \leq \frac{p(p-1)}{2}q,$$

причем равенство достигается только для графов из класса $\mathcal{M}^p(V)$.

Доказательство. Пусть M -граф $M \in \mathcal{M}^{\leq p}(V)$ имеет наибольшее количество ребер среди всех графов из $\mathcal{M}^{\leq p}(V)$. Докажем, что число его ребер равно $\frac{p(p-1)}{2}q$. Предположим, что граф M содержит компоненту связности мощности меньше p . В таком случае в графе M должна содержаться еще хотя бы одна компонента мощности меньше p . Обозначим эти компоненты связности M_1 и M_2 , пусть их мощности соответственно равны p_1 и p_2 , причем $p_1 \leq p_2 < p$. В графе M переместим вершину v из компоненты M_1 в компоненту M_2 , т. е. удалим все ребра, инцидентные вершине v , и добавим все ребра между вершиной v и всеми вершинами компоненты M_2 ; полученный граф обозначим через $M' = (V, E') \in \mathcal{M}^{\leq p}(V)$. Так как изменения производились только в компонентах связности M_1 и M_2 , то

$$|E'| - |E| = p_2 - (p_1 - 1) = p_2 - p_1 + 1 \geq 1.$$

Следовательно, $|E'| > |E|$, но это противоречит тому, что граф M имеет наибольшее число ребер. Значит, все компоненты графа M имеют мощность p , поэтому число ребер в каждой компоненте равно $\frac{p(p-1)}{2}$, а число компонент равно q . ■

Обозначим через \mathcal{K} класс таких графов $G = (V, E)$, что $|V| = pq$ и $|E| \geq \frac{p(p-1)}{2}q$, где p, q — целые положительные числа и $p \geq 3$.

Лемма 3. Пусть $G = (V, E) \in \mathcal{K}$, $M = (V, E') \in \mathcal{M}^{\leq p}(V)$ — произвольное допустимое решение задачи $\mathbf{A}^{\leq p}$ на графе G . Тогда верно неравенство

$$\rho(G, M) \geq |E| - \frac{p(p-1)}{2}q.$$

При этом равенство имеет место тогда и только тогда, когда $M \in \mathcal{M}^p(V)$ и M — подграф графа G .

Доказательство. Сначала докажем неравенство. По определению расстояния

$$\rho(G, M) = |E \setminus E'| + |E' \setminus E| \geq |E \setminus E'| = |E| - |E \cap E'|.$$

Так как $|E| \geq \frac{p(p-1)}{2}q$, а $|E'| \leq \frac{p(p-1)}{2}q$ (по лемме 2), то

$$|E \cap E'| \leq |E'| \leq \frac{p(p-1)}{2}q.$$

Таким образом,

$$\rho(G, M) \geq |E| - \frac{p(p-1)}{2}q \geq 0.$$

Докажем вторую часть утверждения леммы. Пусть $M = (V, E') \in \mathcal{M}^{\leq p}(V)$, причем $\rho(G, M) = |E| - \frac{p(p-1)}{2}q$. Покажем, что $M \in \mathcal{M}^p(V)$. Компоненты связности графа M обозначим через M_1, M_2, \dots, M_l . Пусть $p_i = |M_i| \leq p$ для всех $i \in \{1, \dots, l\}$, где $l \geq q$.

Заметим, что $|E \cap E'| \leq |E'|$. Рассмотрим расстояние между графами G и M :

$$\rho(G, M) = |E \setminus E'| + |E' \setminus E| = |E| + |E'| - 2|E \cap E'| \geq |E| - |E'|.$$

Если найдется такое i , что $p_i < p$, то $|E'| < \frac{p(p-1)}{2}q$ по лемме 2, поэтому

$$\rho(G, M) > |E| - \frac{p(p-1)}{2}q,$$

что противоречит равенству $\rho(G, M) = |E| - \frac{p(p-1)}{2}q$. Следовательно, $p_i = p$ для любого $i \in \{1, \dots, l\}$.

Докажем, что каждая компонента связности графа M является подграфом графа G . Пусть найдется компонента графа M , не являющаяся подграфом графа G . Значит, $|E \cap E'| < \frac{p(p-1)}{2}q$ и, следовательно,

$$\rho(G, M) \geq |E| - |E \cap E'| > |E| - \frac{p(p-1)}{2}q.$$

И вновь получаем противоречие. Таким образом, если для M верно равенство $\rho(G, M) = |E| - \frac{p(p-1)}{2}q$, то $M \in \mathcal{M}^p(V)$ и каждая компонента связности является подграфом графа G .

Доказательство в обратную сторону очевидно. Если $M = (V, E') \in \mathcal{M}^p(V)$ — подграф графа G , то он имеет ровно q компонент, поэтому

$$\rho(G, M) = |E| - |E'| = |E| - \frac{p(p-1)}{2}q.$$

Лемма 3 доказана. ■

Теорема 3. Задача $\mathbf{A}^{\leq p}$ NP-трудна при любом фиксированном $p \geq 3$.

Доказательство. Рассмотрим следующую вспомогательную задачу.

ЗАДАЧА $\overline{\mathbf{A}}^p$. Дан граф $G = (V, E) \in \mathcal{K}$. Существует ли такой M -граф $M \in \mathcal{M}^{\leq p}(V)$, что $\rho(G, M) = |E| - \frac{p(p-1)}{2}q$?

Из леммы 3 следует, что в случае утвердительного ответа на вопрос, поставленный в задаче $\overline{\mathbf{A}}^p$, граф M принадлежит множеству $\mathcal{M}^p(V)$.

Следующая NP -полная задача распознавания содержится в работе [8] под номером ТГ12.

РАЗБИЕНИЕ НА ИЗОМОРФНЫЕ ПОДГРАФЫ. Заданы графы $G = (V, E)$ и $H = (V', E')$, такие, что для некоторого целого числа q выполнено равенство $|V| = q|V'|$. Можно ли разбить вершины графа G на q непересекающихся множеств V_1, V_2, \dots, V_q , таких, что при $1 \leq i \leq q$ подграфы графа G , индуцированные множествами V_i , изоморфны графу H ?

Известно, что эта задача остается NP -полной для любого фиксированного графа H , содержащего по крайней мере 3 вершины.

Используя лемму 3, получаем следующее утверждение:

При $p \geq 3$ задачи $\overline{\mathbf{A}}^p$ и РАЗБИЕНИЕ НА ИЗОМОРФНЫЕ ПОДГРАФЫ, когда в качестве графа H берется полный p -вершинный граф, эквивалентны на классе графов \mathcal{K} . Значит, задача $\overline{\mathbf{A}}^p$ NP -полна при любом фиксированном $p \geq 3$. Задача $\overline{\mathbf{A}}^p$ сводится по Тьюрингу к задаче $\mathbf{A}^{\leq p}$ на графах из \mathcal{K} , поэтому задача $\mathbf{A}^{\leq p}$ при $p \geq 3$ NP -трудна уже на классе графов \mathcal{K} . ■

Следствием из леммы 3 и доказательства теоремы 3 является следующее утверждение.

Следствие 2. Задача \mathbf{A}^p NP -трудна при любом фиксированном $p \geq 3$.

Доказательство. Как уже отмечалось, в случае утвердительного ответа на вопрос, поставленный в задаче $\overline{\mathbf{A}}^p$, граф M принадлежит множеству $\mathcal{M}^p(V)$. Поэтому задача $\overline{\mathbf{A}}^p$ сводится по Тьюрингу к задаче \mathbf{A}^p на графах класса \mathcal{K} . Отсюда следует, что задача \mathbf{A}^p при $p \geq 3$ NP -трудна. ■

ЛИТЕРАТУРА

1. Ильев В. П., Фридман Г. Ш. К задаче аппроксимации графами с фиксированным числом компонент // Докл. АН СССР. 1982. Т. 264. № 3. С. 533–538.
2. Ляпунов А. А. О строении и эволюции управляющих систем в связи с теорией классификации // Проблемы кибернетики. Вып. 27. М.: Наука, 1973. С. 7–18.
3. Фридман Г. Ш. Исследование одной задачи классификации на графах // Методы моделирования и обработки информации. Новосибирск: Наука, 1976. С. 147–177.
4. Tomescu I. La reduction minimale d'un graphe à une reunion de cliques // Discrete Math. 1974. V. 10. No. 1–2. P. 173–179.
5. Zahn C. Approximating symmetric relations by equivalence relations // J. Soc. Indust. Appl. Math. 1964. V. 12. No. 4. P. 840–847.
6. Агеев А. А., Ильев В. П., Кононов А. В., Талевнин А. С. Вычислительная сложность задачи аппроксимации графов // Дискрет. анализ и исслед. опер. Сер. 1. 2006. Т. 13. № 1. С. 3–15.
7. Edmonds J. Paths, trees, and flowers // Canad. J. Math. 1965. V. 17. No. 3. P. 449–467.
8. Гэри М., Джонсон Д. Вычислительные машины и труднорешаемые задачи. М.: Мир, 1982. 416 с.

**ИНТЕРВАЛЬНАЯ НА ОДНОЙ ДОЛЕ ПРАВИЛЬНАЯ РЕБЕРНАЯ
5-РАСКРАСКА ДВУДОЛЬНОГО ГРАФА¹**

А. М. Магомедов, Т. А. Магомедов

Дагестанский государственный университет, г. Махачкала, Россия

E-mail: magomedtagir1@yandex.ru

Пусть в двудольном графе $G = (X, Y, E)$ степень каждой вершины в X равна 2, наибольшая степень вершины в Y равна 5. Найдены условия существования правильной рёберной 5-раскраски графа G , интервальной на множестве X .

Ключевые слова: *двудольный граф, NP-полнота, рёберная раскраска.*

Введение

В работе приняты обозначения и терминология [1].

Рёберной p -раскраской графа G в цвета $1, 2, \dots, p$ называется сюръекция

$$\varphi: E(G) \rightarrow \{1, 2, \dots, p\}.$$

Количество рёбер i -го цвета, инцидентных вершине v , обозначается $\varphi(v, i)$; говорят, что в вершине v цвет i *представлен* $\varphi(v, i)$ раз. *Присутствие* или *отсутствие* цвета i в вершине v означает, что $\varphi(v, i) > 0$ или $\varphi(v, i) = 0$ соответственно. Если $\varphi(v, i) \leq 1$ для каждого цвета i , то рёберная раскраска φ называется *правильной в вершине v* ; если φ является правильной в каждой вершине $v \in V(G)$, то φ называется *правильной рёберной раскраской* графа G . Наименьшее p , такое, что существует правильная рёберная p -раскраска графа G , принято обозначать $\chi'(G)$. Напомним, что через $\Delta(G)$ обозначается наибольшая степень вершины графа G (если граф G однозначно определяется из контекста, обычно применяется сокращенное обозначение Δ).

В [2] доказано, что для простого графа G значение $\chi'(G)$ равно Δ или $\Delta + 1$. Для графа G без петель $\chi'(G) \leq \Delta + k$, где k — максимальное число параллельных ребер; для непустого однородного графа с нечётным числом вершин $\chi'(G) = \Delta + 1$ [1, с. 191]. Из работ [3, 4] и знаменитой теоремы о четырех красках следует, что для всякого 2-связного планарного кубического графа G выполняется равенство $\chi'(G) = 3$. В общем случае для кубического графа G задача уточнения: $\chi'(G) = 3$ или $\chi'(G) = 4$ — является NP-полной [5] (соответствующая гипотеза была выдвинута ранее в [6]). Теорема Кенига о рёберной раскраске [7, с. 80] утверждает, что для двудольного графа G справедливо равенство $\chi'(G) = \Delta$.

Для заданных целочисленных a и b , $a \leq b$, множество $\{a, a + 1, \dots, b\}$ будем называть *интервалом*. Рёберную раскраску, правильную в вершине $v \in V$, называют *интервальной в вершине v* , если цвета рёбер, инцидентных вершине v , заполняют некоторый интервал. Рёберную раскраску графа G называют *интервальной*, если она является интервальной в каждой вершине графа, и *интервальной на множестве $V' \subseteq V(G)$* , если она является интервальной в каждой вершине из V' .

Задача существования правильной рёберной Δ -раскраски двудольного графа $G = (X, Y, E)$, интервальной на множестве X , рассмотрена в работе [8] в связи с прикладной задачей устранения «окон» преподавателей в расписании учебных занятий;

¹Работа поддержана грантом РФФИ № 2010–1.3.2–111–017–12.

в [8], в частности, показано, что если в графе $G = (X, Y, E)$ степени вершин X «преобладают» (над вершинами Y) в следующем смысле:

$$\min_{x \in X} d_G x \geq \max_{y \in Y} d_G y \quad \text{или} \quad \forall (x, y) \in E (d_G x \geq d_G y),$$

то G допускает правильную рёберную Δ -раскраску, интервальную на X (с дополнительным свойством: цвет 1 представлен в каждой вершине X). Задача представляется более содержательной, когда в определённом смысле «преобладают» степени вершин множества Y ; в [9] доказано, что задача о существовании правильной рёберной Δ -раскраски, интервальной на X , NP-полна, даже если

$$\forall x \in X (d_G x \leq \lceil \Delta/2 \rceil); \quad \forall y \in Y (d_G y = \Delta).$$

Иногда в прикладных задачах требуется, чтобы тот или иной цвет присутствовал лишь в определенных вершинах. Рёберную раскраску графа $G = (X, Y, E)$ будем называть *раскраской с ограничением на цвет k* , если цвет k присутствует лишь в вершинах наибольшей степени. Пусть двудольный граф $G = (X, Y, E)$ задает учебные нагрузки к расписанию: количество ребер, соединяющих вершины $x_i \in X$ и $y_j \in Y$, равно количеству уроков, предписанных учителю x_i провести в классе y_j . Тогда, рассматривая цвета $1, 2, \dots, \Delta$ в качестве номеров академических часов, получим, что требование планировать урок учителю x_i в академический час с номером Δ лишь в том случае, когда учебная нагрузка учителя x_i содержит Δ уроков, равносильно условию ограничения на цвет Δ .

Двудольный граф $G = (X, Y, E)$, в котором $d_G x = a$ для всех $x \in X$, будем называть (a, b) -графом, (a, \hat{b}) -графом, (a, \bar{b}) -графом или (a) -графом, если соответственно $d_G y = b$ для всех $y \in Y$; $d_G y \leq b$ для всех $y \in Y$; $\max_{y \in Y} d_G y = b$ или значения степеней вершин Y безразличны.

Докажем, что задача о существовании у заданного $(2, \bar{5})$ -графа $G = (X, Y, E)$ правильной рёберной 5-раскраски, интервальной на X и с ограничением на произвольно выбранный цвет $k = 1, 3$ или 5, разрешима за полиномиальное время. Известно, что задача о существовании интервальной Δ -раскраски у двудольного графа G с $\Delta = 5$ NP-полна [10].

Отметим также следующие результаты. Впервые вопрос об интервальной раскрашиваемости двудольных графов, видимо, сформулировал Н. М. Hansen [11]. Он, в частности, доказал, что в следующих случаях двудольный граф G интервально раскрашиваем: 1) G — полный граф, 2) G является $(2, 2b)$ -графом, 3) степени вершин в графе G не превышают трёх. Если G является $(2, \Delta)$ -графом, то при чётном Δ существует рёберная интервальная Δ -раскраска [12], а при нечётном $\Delta \geq 3$ — рёберная интервальная раскраска в Δ или $(\Delta + 1)$ цветов [13]. В [14] доказана NP-полнота задачи о существовании интервальной Δ -раскраски для $(4, 3)$ -графа (соответствующая гипотеза была сформулирована в [15]); NP-полнота аналогичной задачи для $(6, 3)$ -графа установлена в [16].

Представляют интерес контрпримеры, уточняющие значения Δ и n (количество вершин), для которых можно указать графы, не обладающие рёберной интервальной раскраской: $\Delta = 14$ и $n = 21$ (С. В. Севастьянов); $\Delta = 13$ и $n = 27$ (Р. Erdős); $\Delta = 15$ и $n = 19$ (М. Malafiejski). С привлечением компьютерных вычислений К. Giago установил, что при $n \leq 14$ любой двудольный граф допускает рёберную интервальную раскраску.

1. Вспомогательные утверждения

Для $(2, \bar{5})$ -графа $G = (X, Y, E)$ примем следующие определения:

- 1) множество $E' \subseteq E$, такое, что порожденный на E' подграф графа G является $(1, \bar{2})$ -графом, называется *рёберным 2-каркасом* графа G ;
- 2) рёберная 2-раскраска φ графа G называется *выровненной*, если выполнены следующие условия:
 - а) $\varphi(x, 1) = \varphi(x, 2) = 1$ для всех $x \in X$;
 - б) для каждой вершины $y \in Y$, такой, что $d_{Gy} \geq 3$, справедливо неравенство $|\varphi(y, 1) - \varphi(y, 2)| \leq 1$; другими словами, для любого $i = 1, 2$ $\varphi(y, i) \leq 3$ для всех $y \in Y$ и равенство $\varphi(y, i) = 3$ возможно тогда и только тогда, когда $d_{Gy} = 5$, что равносильно $\varphi(y, 3 - i) = 2$;
- 3) набор (φ, E') из рёберной 2-раскраски φ и рёберного 2-каркаса E' графа G будем называть *2-каркас-раскраской*, если ни одной вершине из множества Y не инцидентны два ребра одного цвета из множества E' .

Далее понадобятся следующие обозначения для графа $G = (V, E)$: $\Gamma(S)$ — множество рёбер, смежных вершинам $S \subseteq V$; $\text{density}(G) = |E|/|V|$.

В [17] доказано следующее утверждение.

Утверждение 1. Пусть $M(v, e)$ — время, требуемое для вычисления минимального разреза сети с v вершинами и e ребрами. Тогда в графе, состоящем из n вершин и m ребер, подграф G' с наибольшим значением $\text{density}(G')$ может быть найден за время $O(M(n, n + m) \log n)$.

Утверждение 2. Если $(2, \bar{5})$ -граф $G = (X, Y, E)$ обладает правильной рёберной 5-раскраской φ , интервальной на X , то

$$\forall S \subseteq X \quad |S| \leq 2|\Gamma(S)|. \quad (1)$$

Доказательство. Пусть $S \subseteq X$; $G_s = (S, \Gamma(S), E_s)$ — подграф графа G , порожденный множеством S . Поскольку раскраска φ является интервальной на S , то цвета рёбер, инцидентных произвольной вершине $x \in S$, образуют одну из пар

$$1 \text{ и } 2; \quad 2 \text{ и } 3; \quad 3 \text{ и } 4 \quad \text{или} \quad 4 \text{ и } 5.$$

Таким образом, в вершине $x \in S$ представлен один и только один из цветов 2 и 4. Тогда можно указать цвет $i_0 \in \{2, 4\}$, который представлен хотя бы в половине вершин из S ; следовательно, не менее $|S|/2$ рёбер множества E_s закрашены в цвет i_0 . Отсюда и из правильности раскраски φ следует, что цвет i_0 представлен в не менее чем $|S|/2$ вершинах множества $\Gamma(S)$. Следовательно, $|S|/2 \leq |\Gamma(S)|$. ■

Далее покажем, что условия (1) являются и достаточными для существования в $(2, \bar{5})$ -графе $G = (X, Y, E)$ правильной рёберной 5-раскраски φ , интервальной на X . Отметим аналогию условий (1) с условиями теоремы Холла [1, с. 164] о существовании в двудольном графе $G = (X, Y, E)$ полного паросочетания множества X с множеством Y . Естественно, следует рассмотреть вопрос эффективной проверки условий (1).

Утверждение 3. Проверка условий (1) выполнима за полиномиальное время.

Доказательство. Заменяем в графе $G = (X, Y, E)$ каждую вершину $x \in X$ и рёбра (x, y') и (x, y'') одним ребром (y', y'') ; полученный граф обозначим через $H = (Y, Z)$, где Z — множество рёбер. Условия (1), очевидно, равносильны следующим: для любого $Z' \subseteq Z$ и подграфа $H' = (Y', Z')$ графа H , порожденного множеством Z' , выполняется неравенство $|Z'|/|Y'| \leq 2$.

Воспользуемся утверждением 1: найдем в графе H подграф G' с наибольшим значением $\text{density}(G')$ и заметим, что проверка условий (1) сводится к проверке неравенства $\text{density}(G') \leq 2$. ■

Результат работы [18] удобно привести в следующем виде.

Утверждение 4. Если $(2, \bar{5})$ -граф $G = (X, Y, E)$ удовлетворяет условиям (1), то G обладает рёберным 2-каркасом.

Нам понадобится также следующее утверждение из [19].

Утверждение 5. Если в связном (2) -графе G существует некоторый 2-каркас, то граф G обладает 2-каркас-раскраской.

Утверждение 6. Если связный (2) -граф $G = (X, Y, E)$ обладает полным паросочетанием X с Y , то G содержит не более одного цикла.

Доказательство. Пусть G содержит два различных цикла: C_1 и C_2 . Обозначим через $G' = (X', Y', E')$ минимальный по числу рёбер связный подграф графа G , включающий C_1 и C_2 . Тогда $\Gamma(X') = Y'$; $d_{G'}x = 2$ для всех $x \in X$; $d_{G'}y \geq 2$ для всех $y \in Y'$; для некоторого $y \in Y'$ выполняется условие $d_{G'}y > 2$.

Отсюда получаем $|X'| > |\Gamma(X')|$. Но это противоречит упомянутой выше теореме Холла, согласно которой $G = (X, Y, E)$ содержит полное паросочетание X с Y тогда и только тогда, когда для любого $X' \subseteq X$ выполняется неравенство $|X'| \leq |\Gamma(X')|$. ■

Следствие 1. Если $G = (X, Y, E)$ — связный (2) -граф, обладающий полным паросочетанием X с Y , то G — либо дерево, либо цикл, либо состоит из простого цикла и набора деревьев, «прикреплённых» к циклу в некоторых вершинах из Y .

Утверждение 7. Пусть $(2, \bar{5})$ -граф $G = (X, Y, E)$ обладает 2-каркас-раскраской (φ, E') . Тогда каждый подграф $G_i = (X_i, Y_i, E_i)$, $i = 1, 2$, графа G , порожденный множеством $E_i = \{e \in E : \varphi(e) = i\}$, допускает интервальную раскраску φ' в три цвета — $2i - 1$, $2i$ и $2i + 1$, такую, что в вершине $y \in Y_i$ цвет 3 представлен тогда и только тогда, когда $d_{G_i}y = 3$.

Доказательство. По определению 2-каркас-раскраски (φ, E') , каждый из подграфов $G_i = (X_i, Y_i, E_i)$ является $(2, \bar{3})$ -графом, который обладает полным паросочетанием X_i с Y_i . Исходная раскраска φ более не потребуется; удалим цвета всех рёбер.

Покажем сначала, что G_1 допускает интервальную раскраску в три цвета — 1, 2 и 3, такую, что цвет 3 представлен только в вершине $y \in Y$ степени 3. Вершины множества X_1 (Y_1) будем называть « X_1 -вершинами» (соответственно « Y_1 -вершинами»).

Рассмотрим случай, когда граф G_1 содержит некоторый цикл C . Выберем одно из направлений обхода цикла C и, начиная с произвольной Y_1 -вершины цикла, выполним обход, поочередно закрашивая рёбра C в цвета 1, 2, 1, 2, ..., 1, 2 (любой цикл в двудольном графе имеет чётную длину).

Если граф G_1 состоит только из цикла C , то закрашивание завершено.

Пусть G_1 , кроме цикла C , содержит (см. следствие 1) и некоторые деревья, «прикреплённые» к C . Для каждой Y_1 -вершины y цикла C , к которой «прикреплено» некоторое дерево F , выполним следующую процедуру.

Процедура закрашивания дерева. Сориентируем рёбра F так, чтобы получить ориентированное корневое дерево F' с корнем в вершине y , и закрасим дуги дерева F' в цвета множества $\{1, 2, 3\}$ по следующему правилу:

- 1) дугу с началом в корневой вершине y закрасим в цвет 3; теперь в вершине y представлены цвета 1, 2 и 3;

- 2) для каждой Y_1 -вершины y' дерева F' закрасим дугу с концом в y' в цвет 2; теперь в каждой Y_1 -вершине дерева (независимо от степени вершины) точно один раз представлен цвет 2;
- 3) **если** из некорневой Y_1 -вершины y' дерева F' выходят две дуги, **то** одну из них закрасим в цвет 1, другую — в цвет 3; теперь в каждой Y_1 -вершине дерева, степень которой равна 3, представлены цвета 1, 2 и 3; **иначе** закрасим дугу с началом в y' в цвет 1; теперь в каждой Y_1 -вершине степени 2 представлены цвета 1 и 2.

Конец процедуры.

Если G_1 состоит из единственного дерева F , то выберем любую его висячую вершину y и выполним описанную выше процедуру закрашивания дерева с заменой п. 1 на следующий: «дугу с началом в y закрасим в цвет 1».

Полученную рёберную раскраску графа G_1 в цвета 1, 2 и 3 обозначим через φ' и подытожим свойства φ' :

- 1) в каждой X_1 -вершине представлен цвет 2 и один из цветов 1 или 3 (следовательно, раскраска φ' интервальна на X_1);
- 2) в каждой Y_1 -вершине y степени 1, 2 и 3 представлены цвета 2; 1 и 2; 1, 2 и 3 соответственно (следовательно, раскраска φ' интервальна на Y_1).

Таким образом, для G_1 требуемая раскраска построена. Аналогично построим раскраску для G_2 , после чего перекрасим в G_2 рёбра цвета 1 в цвет 5, а рёбра цвета 2 — в цвет 4. ■

2. Условия существования правильной рёберной 5-раскраски, интервальной на X

Теорема 1. Пусть $\Delta = 5$; k — произвольный элемент множества $\{1, 3, 5\}$. Тогда $(2, \overline{\Delta})$ -граф $G = (X, Y, E)$ обладает правильной рёберной Δ -раскраской, интервальной на X и с ограничением на цвет k , если и только если выполнены условия (1).

Доказательство. Необходимость доказана в утверждении 2.

Достаточность. Пусть (1) выполнено и, согласно утверждениям 4, 5 и 7, для графа G существует рёберная раскраска φ и разбиение на два $(2, \hat{3})$ -графа $G_1 = (X_1, Y_1, E_1)$ и $G_2 = (X_2, Y_2, E_2)$, таких, что: 1) $d_{G_i}y = 3$ тогда и только тогда, когда $d_{G_{3-i}}y = 2$; 2) сужение φ' раскраски φ на G_1 является интервальной раскраской графа G_1 в цвета 1, 2, 3, такой, что в Y_1 -вершине y цвет 3 присутствует тогда и только тогда, когда $d_{G_1}y = 3$; 3) сужение φ' раскраски φ на G_2 является интервальной раскраской графа G_2 в цвета 3, 4, 5, такой, что в Y_2 -вершине y цвет 3 представлен тогда и только тогда, когда $d_{G_2}y = 3$.

Поэтому в вершине $y \in Y$ цвет 3 представлен тогда и только тогда, когда $d_G y = 5$, и точно один раз. Поэтому φ является правильной рёберной 5-раскраской графа G в цвета множества $\{1, 2, 3, 4, 5\}$, интервальной на X . Ясно, что в вершине $y \in Y$ цвет 3 представлен в том и только в том случае, когда $d_G y = 5$. Для $k = 3$ теорема доказана.

Ввиду симметрии случаев $k = 1$ и 5 ограничимся доказательством для случая $k = 5$. Покажем, как преобразовать φ в правильную рёберную 5-раскраску в цвета множества $\{1, 2, 3, 4, 5\}$, интервальную на X и такую, что цвет 5 представлен в вершине $y \in Y$ тогда и только тогда, когда $d_G y = 5$.

Пусть $y \in Y$, $\varphi(y, 5) = 1$, $d_G y < 5$.

В цвет 5 закрашены только рёбра графа G_2 , пусть ребро $e = (x, y) \in E_2$ имеет цвет 5; другое ребро из E_2 , инцидентное вершине x , обозначим (x, y') . Так как $d_G y < 5$, то $d_{G_1} y < 3$ и $d_{G_2} y < 3$; поэтому $\varphi(y, 3) = 0$. Перекрасим ребро (x, y) в цвет 3.

Во-первых, цвета ребер, инцидентных вершине y , различны и после перекраски ребра (x, y) . Во-вторых, поскольку до перекраски цвета ребёр, инцидентных вершине x , образовывали интервал и цвет ребра (x, y) был равен 5, то цвет ребра (x, y') равен 4. Поэтому после перекраски ребра (x, y) в вершине x присутствуют цвета 3 и 4, следовательно, раскраска остается интервальной в вершине x . В вершине y , $d_G y < 5$, цвет 5 теперь отсутствует; таким образом, количество вершин множества Y , степени которых меньше 5 и в которых присутствовал цвет 5, уменьшилось на единицу. ■

Заключение

Доказано, что правильная рёберная раскраска $(2, \bar{5})$ -графа $G = (X, Y, E)$, интервальная на X и с ограничением на любой из цветов множества $\{1, 3, 5\}$, существует тогда и только тогда, когда выполнены условия (1). Из утверждений 1 и 3 видно, что проверка условий (1) достигается за полиномиальное время.

ЛИТЕРАТУРА

1. *Свами М., Тхуласираман К.* Графы, сети и алгоритмы. М.: Мир, 1984.
2. *Визинг В. Г.* Об оценке хроматического класса p -графа // Дискретный анализ. Новосибирск: Институт математики СО АН СССР, 1964. Вып. 3. С. 25–30.
3. *Tait P. G.* Remarks on the previous communication // Proc. Roy. Soc. Edin. 1880. V. 10. P. 729.
4. *Tait P. G.* Note on a theorem in the geometry of position // Trans. Roy. Soc. Edin. 1880. V. 29. P. 657–660.
5. *Holyer J.* The NP-completeness of edge-coloring // Siam J. Comput. 1981. V. 10. No. 4. P. 718–720.
6. *Garey M. R. and Johnson D. S.* Computers and Intractability. San Francisco: W. H. Freeman and Company, 1979.
7. *Ловас Л., Пламмер М.* Прикладные задачи теории графов. Теория паросочетаний в математике, физике, химии: пер. с англ. М.: Мир, 1998.
8. *Асратян А. С., Камалян Р. Р.* Интервальные раскраски рёбер мультиграфа // Прикладная математика. Ереван: Изд-во Ереван. ун-та, 1987. Вып. 5. С. 25–34.
9. *Магомедов А. М.* Непрерывное расписание для специализированных процессоров без отношения предшествования // Вестник Московского энергетического института. 2009. № 5. С. 14–17.
10. *Giaco K.* The complexity of consecutive Δ -coloring of bipartite graphs: 4 is easy, 5 is hard // Ars Combin. 1997. V. 47. P. 287–298.
11. *Hansen H. M.* Scheduling with minimum waiting periods (in Danish) // Master Thesis. Odense University. Odense, Denmark, 1992.
12. *Магомедов А. М., Рашида А.* Матрица расписания с двумя ненулевыми элементами в строке // Вестник Дагестанского госуниверситета. 1999. Вып. 4. С. 12–15.
13. *Hanson D., Loten C. O. M., and Toft B.* On interval colourings of bi-regular bipartite graphs // Ars Combinat. 1998. V. 4. P. 23–32.
14. *Pyatkin A. V.* Interval coloring of $(3,4)$ -biregular bipartite graphs having large cubic subgraphs // J. Graph Theory. 2004. V. 47. No. 2. P. 122–128.
15. *Jensen T.R. and Toft B.* Graph coloring problems. New York: Wiley-Interscience series in discrete mathematics and optimization, 1995.
16. *Asratian A. S. and Casselgren C. J.* Some results on interval edge colorings of (α, β) -biregular bipartite graphs // Department of Mathematics, Linköping University S-581 83. Linköping, Sweden, 2007.

17. *Goldberg A. V.* Finding a maximum density subgraph. Berkeley: University of California / Technical Report UCB/CSD 84/171, CA. 1984.
18. *Магомедов А. М.* Два частичных паросочетания в двудольном графе специального вида // Материалы X Междунар. семинара «Дискретная математика и ее приложения» (Москва, МГУ, 1–6 февраля 2010 г.) / под ред. О. М. Касим-Заде. М.: Изд-во мехмата МГУ, 2010. С. 312–313.
19. *Магомедов А. М.* Об одной специальной рёберной 2-раскраске // Научно-технические ведомости СПбГПУ. Сер. Информатика. Телекоммуникации. Управление. Раздел «Математическое моделирование: методы, алгоритмы, технологии». 2011. № 2(120). С. 156–159.

**СТРУКТУРНЫЕ И КОММУНИКАТИВНЫЕ СВОЙСТВА
ЦИРКУЛЯНТНЫХ СЕТЕЙ**

Э. А. Монахова

*Институт вычислительной математики и математической геофизики СО РАН,
г. Новосибирск, Россия***E-mail:** emilia@rav.sccc.ru

Циркулянтные сети интенсивно исследуются последние 30 лет и находят широкое применение в различных областях информатики и дискретной математики. По ним два обзора было опубликовано на английском языке (Дж.-К. Бермонда, Ф. Комелласа и Д. Ф. Хсу в 1995 г. и Ф. К. Хванга в 2003 г.) и один на русском языке (О. Г. Монахова и Э. А. Монаховой, 2000 г.). Настоящий обзор дополнительно включает результаты, которые не были отражены в упомянутых источниках, а также новые результаты, полученные в области исследования неориентированных циркулянтных сетей в последние годы.

Ключевые слова: *сети связи, циркулянтные графы, диаметр, маршрутизация, трансляционные и полные обмены.*

Введение

Циркулянтные сети (графы) и их разнообразные приложения являются объектом интенсивных исследований в информатике и дискретной математике (см., например, [33, 38, 40, 65, 75, 81, 88, 89, 103, 105, 111]). Они находят приложение в проектировании вычислительных сетей, сетей передачи данных и распределенных вычислениях. Циркулянтные графы реализованы как коммуникационные сети в параллельных вычислительных системах ILLIAC-IV, MPP, Intel Paragon, Cray T3D, SONET, отечественной системе МИКРОС и др., использовались также в качестве структур многомодульной высокоскоростной памяти вычислительных систем [106]. В настоящее время расширяются возможности практического применения циркулянтных сетей и их обобщений — как основы структуры в мультипроцессорных кластерных системах [86], в модели «малого мира» (small-world networks) [45], оптических сетях [91], моделях химических реакций [29], клеточных нейронных сетях [23], что объясняется высокими показателями надежности, наращиваемости, модульности и связности этих графов. Важным приложением двумерных и трехмерных циркулянтных графов является их использование в теории кодирования при построении совершенных кодов, исправляющих ошибки [11, 78].

Определение 1. Пусть s_1, s_2, \dots, s_k, n — целые числа, такие, что $1 \leq s_1 < s_2 < \dots < s_k < n$. Неориентированный граф G с множеством вершин $V = \{0, 1, \dots, n-1\}$ и множеством ребер $E = \{(i, j) : |i - j| \equiv s_m \pmod{n}, m = 1, \dots, k\}$, называется *циркулянтной сетью*.

Другими словами, циркулянтный граф есть граф Кэли, чья матрица смежности есть циркулянт. Элементы порождающего множества $S = \{s_1, s_2, \dots, s_k\}$ называются *образующими* (хордами). Параметрическое описание вида $(n; S)$ полностью определяет циркулянт порядка n и *размерности* k . Степень циркулянта равна $2k$, если $s_k \neq n/2$. Если n четное и $s_k = n/2$, то циркулянт имеет степень $2k - 1$. Каждое $s_i \in S$, взаимно

простое с n , влечет гамильтонов цикл в графе. Примеры циркулянтов размерностей 2 и 3 показаны на рис. 1.

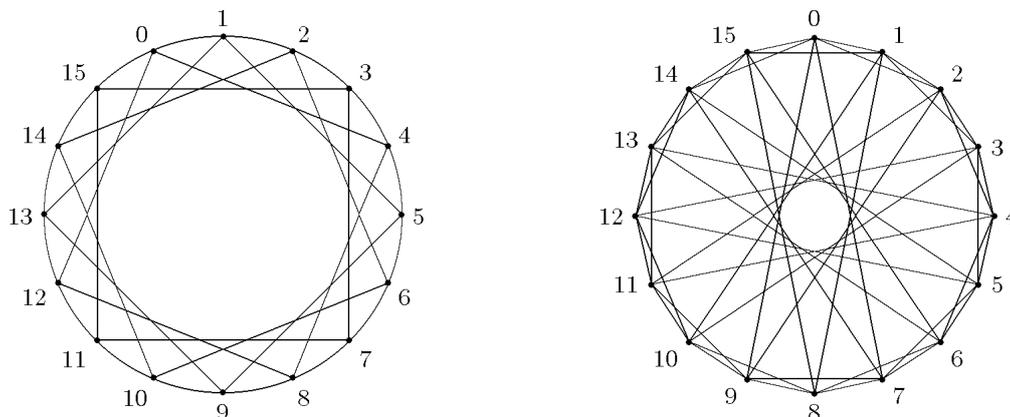


Рис. 1. Циркулянты $C(16; 1, 4)$ и $C(16; 1, 2, 7)$

Согласно Ф. Дэвису [47], циркулянты как математический объект введены Э. Каталаном в 1846 г. Циркулянтные графы также известны в зарубежной литературе как звездные многоугольники (star-polygon graphs, point-symmetric graphs) [108], циклические графы (cyclic graphs) [46, 56], распределенные кольцевые сети (distributed loop networks, multi-loop networks) [33, 35, 41, 49, 65, 87, 115], хордальные кольца (chordal rings) [30, 39, 73, 90, 91], multiple fixed step graphs [52, 72], в отечественной литературе — как D_n -графы (диофантовы структуры) [3, 14, 15].

Впервые циркулянтные графы получили пристальное внимание как перспективные сети связи однородных вычислительных систем в [3, 4, 7, 9, 14]. Результаты сравнения циркулянтных сетей с рядом популярных топологий вычислительных систем, в том числе с гиперкубами, показали [21], что они являются лучшей топологией вычислительных систем, чем гиперкубы, по показателям структурной живучести, надежности и связности, а также требуют меньшего числа межпроцессорных обменов при решении вычислительных задач и задач системного управления.

В настоящей работе рассматриваются только неориентированные циркулянты. В литературе изучается также их ориентированный вариант (см., например, обзоры в [33, 60, 64, 65]): в графе $G(n; s_1, s_2, \dots, s_k)$ с множеством вершин $V = \{0, 1, \dots, n - 1\}$ каждая вершина i связана с k вершинами $i + s_1, i + s_2, \dots, i + s_k \pmod{n}$.

1. Основные свойства циркулянтов

Прежде всего нас интересуют только связные циркулянты. Граф называется связным, если существует по крайней мере один путь между любыми двумя его вершинами. В дальнейшем под (a, b) понимается наибольший общий делитель чисел a и b .

В [3, 37] показано, что циркулянтный граф $C(n; S)$ является связным, если и только если $(s_1, s_2, \dots, s_k, n) = 1$. Для $k = 2$ в [34] доказано, что если $(n, s_1, s_2) = 1$, то циркулянт $C(n; s_1, s_2)$ может быть разложен на два гамильтоновых цикла.

Лемма 1 [87]. Если $(n, s_1) = 1$, то в циркулянтном графе $C(n; s_1, s_2)$ существует гамильтонов цикл, использующий только образующую s_1 .

В [37] аналогичное свойство доказано для циркулянтов любой степени. Тем не менее циркулянт может быть связным и иметь гамильтонов цикл, но не иметь гамильтонова цикла, порожденного одной из образующих (например, граф $C(24; 3, 4)$).

При решении проблем синтеза и выбора структур вычислительных систем важным является вопрос об изоморфизме графов. Из симметрии циркулянтов вытекает

Лемма 2 [3, 108]. Циркулянты $C(n; s_1, \dots, s_i, \dots, s_k)$ и $C(n; s_1, \dots, n - s_i, \dots, s_k)$ изоморфны.

Это свойство позволяет ограничиться рассмотрением циркулянтных графов с обрезающими, не превосходящими $\lfloor n/2 \rfloor$.

Лемма 3 [3, 26, 108]. Если $(n, t) = 1$, то $C(n; s_1, s_2, \dots, s_k)$ и $C(n; ts_1, ts_2, \dots, ts_k)$ изоморфны (образующие $ts_i, i = 1, \dots, k$, взяты по модулю n).

Важными метрическими характеристиками графа являются диаметр и среднее расстояние (соответствуют максимальной и средней структурным задержкам в сети).

Определение 2. Диаметр графа C называется $d(n; S) = \max_{i, j \in V} d(i, j)$, где $d(i, j)$ — длина кратчайшего пути между вершинами i и j , принадлежащими C .

При фиксированном k пусть $D(n)$ означает точную нижнюю границу диаметра для любого n . Заметим, что $d_{\min}(n) = \min_S \{d(n; S)\} \geq D(n)$ для любого n . В [14, 41, 49] показано, что существуют n и k , такие, что $d_{\min}(n) > D(n)$.

Определение 3. Средним расстоянием (средним диаметром в некоторых источниках) графа $C(n; S)$ называется $\bar{d}(n; S) = \sum_{i, j} d(i, j) / (n(n - 1))$.

Как показали исследования, наилучшими структурами вычислительных систем по различным критериям функционирования (структурной живучести, надежности, производительности, самодиагностируемости и др.) при одинаковом числе вычислительных модулей и линий связи у каждого модуля являются структуры с минимальными диаметром и средним расстоянием [1, 2, 4, 5, 8, 9, 10, 24, 25].

Фундаментальная проблема теории графов — синтез оптимальных графов — состоит в поиске графов с минимальным диаметром и/или минимальным средним расстоянием среди регулярных графов с заданными степенью и числом вершин.

Рассмотрим верхнюю границу максимально возможного числа вершин в циркулянтах. Для циркулянтного графа размерности k пусть $P'(d, k)$ означает число вершин, которые находятся на расстоянии не более d от вершины 0 , а $P(d, k)$ — верхнюю границу $P'(d, k)$. Пусть $S'(d, k) = P'(d, k) - P'(d - 1, k)$ и $S(d, k)$ — верхняя граница $S'(d, k)$. Значение $P(d, k)$ для циркулянта диаметра d и размерности k получено Ч. К. Вонгом с соавт. [111, 112] и В. В. Корнеевым [9] (для общего случая КАИС-структур [4, 10]). При выводе выражения $P(d, k)$ в [112] рассматривается множество точек в k -мерном Евклидовом пространстве.

Лемма 4 [111]. Пусть $S(m, k) = |\{(x_1, x_2, \dots, x_k) \in \mathbb{Z}^k : \sum_{i=1}^k |x_i| = m\}|$, где $m \geq 0$ — целое. Тогда

$$S(m, k) = \begin{cases} \sum_{i=0}^{k-1} C_k^i C_{m-1}^{k-i-1} 2^{k-i} & \text{для } m \geq 1, \\ S(0, k) = 1. & \end{cases}$$

Теорема 1. Пусть $P(m, k) = |\{(x_1, x_2, \dots, x_k) \in \mathbb{Z}^k : \sum_{i=1}^k |x_i| \leq m\}|$, где $m \geq 0$ — целое. Тогда

$$P(m, k) = \sum_{i=0}^k C_k^i C_m^{k-i} 2^{k-i}.$$

Этот результат также получили Ф. Бош и Дж. Ванг [38] и Ф. П. Муга [84].

В циркулянте размерности k функция $P(m, k)$ определяет максимальное число вершин, которые могут быть достигнуты из любой вершины графа самое большее за m шагов. Приведем выражения экстремальной функции $P(m, k)$ для $1 \leq k \leq 4$:
 $P(m, 1) = 2m + 1$; $P(m, 2) = 2m^2 + 2m + 1$; $P(m, 3) = \frac{4}{3}m^3 + 2m^2 + \frac{8}{3}m + 1$;
 $P(m, 4) = \frac{2}{3}m^4 + \frac{4}{3}m^3 + \frac{10}{3}m^2 + \frac{8}{3}m + 1$.

Далее функцию $P(d, 2) = 2d^2 + 2d + 1$ будем обозначать n_d .

Отметим, что $P(m, k)$, $k \geq 2$, определяет «объем» k -мерного октаэдра, а $S(m, k)$ — «площадь» его поверхности [112].

Ч. К. Вонг и Д. Кошперсмит [111], используя функцию $P(m, k)$, установили нижние границы диаметра и среднего расстояния циркулянтов порядка n и размерности k :

$$d(n; S) \geq d_{\min}(n) \geq \frac{1}{2}(k!n)^{\frac{1}{k}} - \frac{1}{2}(k+1) \approx \frac{1}{2}(k!)^{\frac{1}{k}}n^{\frac{1}{k}},$$

$$\bar{d}(n; S) \geq \bar{d}_{\min}(n) \geq \frac{1}{2} \frac{k}{(k+1)!n} \left((k!n)^{\frac{1}{k}} - (3k+1) \right)^{k+1} \approx \frac{1}{2} \frac{k}{k+1} (k!)^{\frac{1}{k}} n^{\frac{1}{k}},$$

где $\bar{d}_{\min}(n) = \min_S \{\bar{d}(n; S)\}$.

Я. Зеровник и Т. Пизанский [115] описали алгоритм сложности $O(\log n)$ вычисления диаметра графов вида $C(n; s_1, s_2)$. Используя геометрический подход, они свели проблему вычисления диаметра циркулянтов размерности k к эквивалентной проблеме в целочисленной решетке \mathbb{Z}^k .

Для $k = 2$ из [111] следует, что нижняя граница диаметра равна $(\sqrt{2n} - 3)/2$. Точная нижняя граница диаметра двумерных циркулянтов получена в [15]:

$$D(n) = \lceil (\sqrt{2n} - 1 - 1)/2 \rceil.$$

Величина $D(n)$ также найдена в [35, 38, 39, 56].

В [35] показано, что среднее расстояние $\bar{d}(n)$ в случае $k = 2$ асимптотически равно $\sqrt{2n}/3$. В работе [39] дана формула для нижней границы среднего расстояния в циркулянтах размерности $k = 2$:

$$\bar{d}(n) = (n - 1)\sqrt{2n - 1}/3n.$$

Следуя [67], определим экстремальную функцию $M(d, k)$.

Определение 4. Пусть d — положительное целое, $S = \{1, s_2, \dots, s_k\}$ — множество положительных целых, $m(d, S) = \max\{m : d(m; S) \leq d\}$. Для любых заданных целых d и k положим $M(d, k) = \max\{m(d, S) : \exists S (|S| = k)\}$.

Имеем $M(d, k) = P(d, k)$ для $k \leq 2$; $M(d, k) \leq P(d, k)$ для $k > 2$. В 1994 г. Ф. П. Муга [84] получил следующий результат.

Теорема 2. Значение $M(d, k)$ является нечетным числом при любых $d \geq 1$ и $k \geq 1$.

Будем использовать обозначение из [41]: пусть n и k — положительные целые; обозначим через $d(n, k)$ минимально возможное натуральное d , такое, что существует множество образующих $S = \{1, s_2, \dots, s_k\}$, для которого $d(n; S) \leq d$.

В литературе используются различные понятия оптимальности для циркулянтов. Например, в [33, 49] граф $C(n; S)$ называется оптимальным, если $d(n; S) = d_{\min}(n)$, и предельно оптимальным, если $d(n; S) = D(n)$. В настоящей работе будем использовать следующие определения.

Определение 5. Циркулянтный граф $C(n; S)$, величины n и S называются оптимальными, если $d(n; S) = D(n)$, и субоптимальными, если $d(n; S) = D(n) + 1$.

Определение 6. Множество натуральных чисел Θ называется оптимальным (субоптимальным) семейством, если каждое $n \in \Theta$ оптимально (субоптимально).

Определение 7. Циркулянтный граф $C(n; S)$ называется предельно оптимальным, если число вершин $C(n; S)$ на расстоянии m от вершины 0, $m = 0, 1, \dots, D(n) - 1$, равно $S(m, k)$, а на расстоянии $D(n) - (n - P(D(n) - 1, k))$.

Таким образом, предельно оптимальный граф является самым плотным возможным графом для заданных n и k . Каждый предельно оптимальный граф является оптимальным, но не каждый оптимальный — предельным. Предельно оптимальные графы достигают точных нижних границ диаметра и среднего расстояния и соответственно минимумов максимальной и средней структурных задержек, максимумов связности и надежности [9, 38], минимального числа шагов при реализации коммуникационных алгоритмов [8, 21, 80], но существуют не для всех величин n и k [14]. Для важного случая $k = 2$ они существуют для любого числа вершин (см. теорему 3). Следующие проблемы комбинаторной оптимизации для циркулянтных сетей рассматривались многими авторами.

Задача 1. Найти оптимальные сети с минимальным диаметром (возможно, равным $D(n)$) для любого порядка n .

Задача 2. Найти бесконечные семейства циркулянтных графов с максимальным числом вершин (возможно, равным $M(d, k)$) для любого диаметра d .

Такие оптимальные графы, взятые в качестве сетей связи, имеют высокие отказоустойчивость и скорость коммуникаций, минимальную задержку и максимальные связность и надежность [1, 8, 9, 10, 21, 33, 38, 101, 102]. Вторая из проблем относится к проблеме (Δ, d) -графов (в классе циркулянтов), состоящей в максимизации порядка неориентированного регулярного графа степени Δ и диаметра d (Б. Элспас [50]). Следует упомянуть, что одним из первых авторов, кто рассмотрел проблему сокращения диаметра для семейств циркулянтных графов, был Р. С. Вилков [110].

Выше рассматривались только циркулянты с четной степенью. Для полноты описания класса циркулянтных графов представим результат, относящийся к циркулянтам с нечетной степенью. Используя геометрическую модель циркулянтов, Ф. П. Муга [85] получил верхнюю границу $N(d, 2k + 1)$ максимального порядка циркулянтного графа со степенью $2k + 1$ и диаметром d :

$$N(d, 2k + 1) \leq \sum_{i=0}^{k+1} (C_i^k C_i^d 2^i + C_{i-1}^k C_{i-1}^{d-1} 2^{i-1}).$$

Циркулянты с нечетной степенью найдут применение в классе рекурсивных циркулянтов [95].

2. Оптимальные двумерные циркулянты

Случаи, когда $k = 2$ или $k = 3$, широко изучаются благодаря их практическим приложениям.

Для $k = 2$ в 1981 г. доказано [15], что для каждого числа вершин n неориентированные циркулянтные графы могут иметь одновременно и минимальный диаметр $D(n)$, и минимальное среднее расстояние.

Теорема 3 [15]. Пусть $n > 4$ и $D \geq 1$ — целые числа, такие, что $2D^2 - 1 \leq n \leq n_D + 2D + 2$. Тогда циркулянт $C(n; D, D + 1)$ предельно оптимальный.

Отсюда следует интересное свойство: для каждого $D \geq 2$ циркулянты с порядками $n \in \{2D^2 - 1, 2D^2, 2D^2 + 1\}$ имеют два предельно оптимальных описания — $(n; D - 1, D)$ и $(n; D, D + 1)$ (свойство перекрытия описаний). Из теоремы 3 следует

Теорема 4 [15]. Для любого целого $n > 4$ предельно оптимальный двумерный циркулянт порядка n есть

$$C(n; d, d + 1), \text{ где } d = [(\sqrt{2n - 1} - 1)/2], \quad (1)$$

$[x]$ — ближайшее целое к x .

В 1991 г. Р. Байвиде с соавт. [31] получили то же самое семейство циркулянтов (1), но записанное в другом виде:

$$C(n; b - 1, b), \text{ где } b = \lceil \sqrt{n/2} \rceil, \quad n > 2. \quad (2)$$

Дж.-К. Бермонд и соавт. [35] получили интервалы изменения n , сохраняющие свойство перекрытия описаний, на которых циркулянты $C(n; D, D + 1)$ или $C(n; D - 1, D)$ достигают минимума среднего расстояния (заметим, что эти области значений n следуют из теоремы 3). Авторы [35] доказали также, что диаметр графа $C(n; D, D + 1)$ после удаления одной вершины или одного ребра есть самое большее $D + 1$.

В 1985 г. Ф. Бош и Дж. Ванг доказали следующий результат [38]:

Теорема 5. Нижняя граница диаметра $D(n)$, $n > 6$, достигается в циркулянтном графе $C(n; s_1, s_2)$ с образующими $s_1 = D(n)$, $s_2 = D(n) + 1$.

Они также нашли, что эти графы являются оптимальными по отношению ко всем стандартным критериям, относящимся к сетевой надежности, в частности граф $C(n; D(n), D(n) + 1)$ имеет максимальную связность четыре.

В табл. 1 рассмотрены примеры, поясняющие разницу между графами из [15] и [38]. Комбинируя результаты работ [15, 38], представляем в последней колонке расширенные области оптимальности рассмотренных описаний.

Т а б л и ц а 1

Области оптимальности описаний

Семейство графов	min d & min \bar{d} [15]	min d [38]	min d
$C(n; 2, 3)$	$n \in [7, 19]$	$n \in [7, 13]$	$n \in [7, 19]$
$C(n; 3, 4)$	$n \in [17, 33]$	$n \in [14, 25]$	$n \in [14, 33]$
$C(n; 4, 5)$	$n \in [31, 51]$	$n \in [26, 41]$	$n \in [26, 51]$
$C(n; 5, 6)$	$n \in [49, 73]$	$n \in [42, 61]$	$n \in [42, 73]$

Задача исследования найденных оптимальных двумерных циркулянтов как сетей связи компьютерных систем интенсивно изучалась последние 20 лет. Ниже величина $D(n)$ для краткости обозначена через D .

В 1997 г. К. Хубер [63] рассмотрел задачу оптимального проектирования СБИС для оптимальных циркулянтов вида (2) любого порядка n . В 1998 г. А. Л. Листман и др. [72] изучили сетевые свойства двух- и трехмерных циркулянтов, в частности оптимальных графов $C(n_D; D, D + 1)$, и исследовали возможность вложения в них решеток.

В серии работ 1999–2001 гг. [97, 98, 113] и др. авторы рассмотрели оптимальные сети (2) (соответственно (1)) в качестве технической реализации сетей связи вычислительных систем высокой производительности. Они назвали эти сети *midimeu*-сетями.

В. Пуэнте с соавт. [98] показали для них увеличение сетевой производительности при реальных нагрузках, а также уменьшение длины среднего пути сообщения по сравнению с торами. В [97, 98] для этой топологии представлено практическое решение проблемы предотвращения дедлоков (блокировок пути при передаче пакетов). Ю. Янг и соавт. [113] использовали несколько примеров таких графов как базис при проектировании сетей связи массово параллельных систем.

В 2008 г. К. Мартинес с соавт. [11, 78] применили циркулянтные сети размерностей 2 и 3, в частности семейство вида $C(n_D; D, D + 1)$, в теории кодирования при построении совершенных групповых кодов.

В 2009 г. Б. Б. Нестеренко и М. А. Новотарский [23] применили квадратную структуру оптимальных циркулянтов вида

$$C(n; \sqrt{n}/2 - 1, \sqrt{n}), \text{ где } n = 2^{2^i}, i \geq 3,$$

полученных в [31], в качестве базовой структуры дискретных клеточных нейронных сетей. Такие сети ориентированы на моделирование сложных физических процессов (алгоритмы обработки изображений, распознавания образов, оценки динамики механических систем и др.) путем численного решения уравнений математической физики.

Важной характеристикой в проектировании параллельных систем является *расширяемость* (или наращиваемость), т. е. возможность увеличения системной мощности добавлением большего количества вершин в сети без изменения основных связей, свойств и характеристик топологии [9, 16, 68]. Из теоремы 3 следует, что параметры описаний оптимальных графов сохраняются при изменении числа вершин в больших диапазонах. В [16] представлен алгоритм статической реконфигурации циркулянтной сети в вычислительной системе, когда новые модули добавляются в систему. Стоимость реконфигурации с ростом числа вершин n приближается к нулю как $1/\sqrt{n}$.

3. Эквивалентность циркулянтов

Изучение изоморфизма графов является необходимой задачей для решения проблемы синтеза и выбора структур вычислительных систем. Проблемой изоморфизма циркулянтных графов занимались Б. Элспас и Дж. Тернер [51, 108], В. А. Воробьев [3], Ф. Гобел и Е. Нейтел [56], К. Делорме и М. Махео [48], С. А. Евдокимов и И. Н. Пономаренко [6], Б. Манс, Ф. Паппаларди и И. Шпарлинский [73, 74, 75], М. Музычек [88], Х. Фенг и М. Ху [53] и многие другие. Обзор по изоморфизму конечных графов Кэли, где в том числе представлены циркулянтные графы, дан в [70]. Здесь коротко рассмотрим свойства эквивалентности (изоморфизма) циркулянтных графов.

Пусть задано некоторое описание циркулянтного графа C (оптимального или нет) $(n; S) = (n; s_1, \dots, s_i, \dots, s_k)$. Необходимо получить другие его описания. Умножим все $s_i \in S$, $i = 1, 2, \dots, k$, на элемент t приведенной системы вычетов по модулю n . В качестве новых образующих s'_i возьмем остатки от деления ts_i на n , если они не больше $\lfloor n/2 \rfloor$, или дополнения этих остатков до n в противном случае. Назовем данное преобразование, переводящее все s_i в s'_i , эквивалентным преобразованием [51], а отношение между множествами S и S' , а также графами $C(n; S)$ и $C(n; S')$ отношением эквивалентности. Действительно, оно рефлексивно, симметрично и транзитивно. Все графы $C(n; S')$, получаемые из $C(n; S)$, когда t пробегает приведенную систему вычетов по модулю n , образуют класс эквивалентности.

Очевидно, эквивалентность графов влечет их изоморфизм [26, 51]. Например, графы $C(23; 1, 9)$ и $C(23; 1, 5)$ изоморфны, поскольку эквивалентны ($t = 5$). А. Адам [26]

предположил, что обратное также имеет место: любые два изоморфных циркулянтных графа эквивалентны. Оказалось, что в общем случае это неверно. Например, циркулянты $C(16; 1, 2, 7)$ и $C(16; 2, 3, 5)$ изоморфны, но не эквивалентны [51]. Тем не менее в [48, 53] доказано важное свойство: для двумерных циркулянтов общего вида понятия эквивалентности и изоморфизма совпадают при любом порядке графа. Отметим, что в [56] это свойство доказано для двумерных циркулянтов с единичной образующей. Понятия эквивалентности и изоморфизма совпадают для циркулянтов с простым порядком или порядком, равным произведению двух различных простых чисел [37, 108], а также в случаях некоторых других n [37, 74, 88]. В 2003 г. С. А. Евдокимов и И. Н. Пономаренко [6] построили алгоритм полиномиальной сложности для распознавания и проверки изоморфизма произвольных циркулянтных графов.

Чтобы получить все эквивалентные описания циркулянта порядка n , достаточно взять в качестве множителя t те элементы приведенной системы вычетов по модулю n , которые не превосходят $\lfloor n/2 \rfloor$ [14].

Пример 1. Оптимальный двумерный граф с $n = 36$ в силу теоремы 4 имеет образующие $s_1 = 4$ и $s_2 = 5$. Приведенная система вычетов по модулю 36 есть $\{1, 5, 7, 11, 13, 17\}$. Перевод образующих 4 и 5 в s'_1, s'_2 дает новый эквивалентный (изоморфный) оптимальный граф $C(36; 1, 8)$ с единичной образующей.

К сожалению, данный метод не всегда может использоваться для нахождения оптимальных двумерных графов с единичной образующей, так как для некоторых значений n они не существуют.

4. Двумерные циркулянты с единичной образующей

Двумерные циркулянтные графы с единичной образующей $s_1 = 1$ (или кольцевые циркулянты) являются популярной моделью коммуникаций в локальных сетях и архитектурах параллельной обработки. Большое число работ посвящено изучению проблем, связанных с этими сетями: существованию оптимальных графов, изучению диаметра и других структурных характеристик, конструированию алгоритмов обменов для них и др.

В [49] авторы поставили следующую задачу.

Задача 3. Классифицировать все значения n , для которых предельно оптимальные (оптимальные) двумерные кольцевые циркулянты порядка n могут быть найдены.

Во многих работах, изучающих эту проблему, используется следующая геометрическая модель циркулянтов [111] (см. также [14, 109, 114]). Циркулянтный граф $C(n; s_1, s_2)$ конструируется на плоскости Евклида \mathbb{Z}^2 как ромбоподобная конфигурация, если каждую точку решетки (i, j) пометить числом $m \equiv s_1 i + s_2 j \pmod{n}$, где $0 \leq m < n$ — номер вершины графа. В результате все отметки $0 \leq m < n$ повторятся на плоскости бесконечно много раз, образуя укладку (tessellation) \mathbb{Z}^2 . Пример укладки плоскости посредством геометрической модели графа $C(25; 1, 7)$ изображен на рис. 2. Расширение геометрической модели циркулянтных графов для $k \geq 3$ очевидно.

Обозначим $R(d) = \{n_{d-1} + 1, \dots, n_d\}$, $d > 0$. Таким образом, натуральный ряд чисел делится на интервалы $R(d)$, $d > 0$, и $|R(d)| = 4d$. Например, $R(1) = \{2, \dots, 5\}$, $R(2) = \{6, \dots, 13\}$, $R(3) = \{14, \dots, 25\}$. В областях $R(d)$ линии значений $n = q_1[d] = 2d^2 - d$, $n = q_2[d] = 2d^2$, $n = q_3[d] = 2d^2 + d$ и $n = n_d$, $d \geq 1$, играют важную роль в конструировании оптимальных семейств. Эти значения n изучались в ряде работ, где доказано существование предельно оптимальных графов $C(n; 1, s)$ для них. Точки

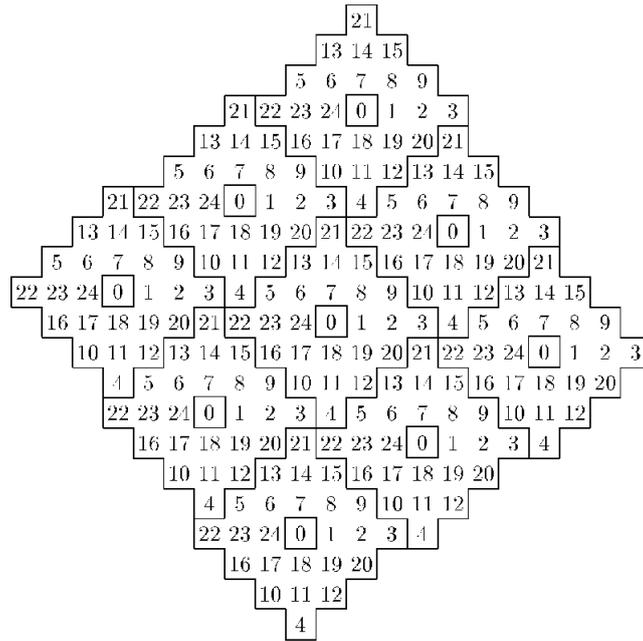


Рис. 2. Пример укладки плоскости

$n = q_1[d], q_2[d], q_3[d]$ также имеют много неэквивалентных оптимальных описаний вида $(n; 1, s)$.

В работах [14, 35, 36, 43, 44, 49, 61, 62, 79, 109] рассмотрены различные подходы к решению задачи 3 и получены бесконечные семейства графов с аналитическим описанием.

Д. Ду и соавт. [49] дали следующую верхнюю границу диаметра графа $C(n; 1, s)$:

$$d(n; 1, s) \leq \max\{\lfloor n/s \rfloor + 1, n - \lfloor n/s \rfloor s - 2, (\lfloor n/s \rfloor + 1)s - n - 1\}$$

и нашли 16 бесконечных семейств оптимальных графов $C(n; 1, s)$, для которых точная нижняя граница диаметра достигается. Они использовали следующую лемму для получения параметров оптимальных (и субоптимальных) графов.

Лемма 5 [49]. Пусть $n = 2d^2 + qd + h$, где $2 \leq q \leq 6$ и $0 \leq h < d$. Тогда $d(n; 1, 2d + (q - 1)) \leq \max\{d + 1, d + h - 2, d + q - h - 2\}$.

В работе [49] (см. также [109] и теорему 8 из [79]) доказана следующая

Теорема 6. Пусть $n = n_d - 1$, $d \geq 1$. Тогда $d_{\min}(n) = d + 1$.

Ф. Гобел и Е. Нейтел [56] получили некоторые результаты по исследованию хромотического числа и линейной транзитивности двумерных кольцевых сетей. Они нашли следующие зависимости для диаметра графов $C(n; 1, s)$ с заданной образующей s :

Теорема 7. Если $d(n; 1, s) \geq n/s$, то $d(n + 2s; 1, s) = d(n; 1, s) + 1$.

Теорема 8. Для любого $M > 0$ существуют n и s , такие, что $d(n + 2s; 1, s) > d(n; 1, s) + M$.

Д. Цвиели [109] описала бесконечные семейства графов $C(n; 1, s)$ с заданной образующей s и их диаметры.

Лемма 6. Пусть $l \geq 1$, $m \geq -l + 1$. Тогда $d(n; 1, s) = l + \lceil m/2 \rceil$, где $s = 2l + 1$, $n = n_l + ms$ или $s = 2l$, $n = 2l^2 + l + ms$.

Часть графов из описанных семейств являются оптимальными. Используя лемму 6, а также геометрические трансформации циркулянтов на плоскости, Д. Цвиели нашла

четыре бесконечных семейства оптимальных величин n . Они включают все семейства из [49], но, в отличие от [49], содержат $O(\sqrt{d})$ величин n в каждом $R(d)$. В частности, для точек $n = q_i[d]$, $i = 1, 2, 3$, получена

Теорема 9. Пусть $n = q_i[d]$, $i = 1, 2, 3$. Образующие оптимальных графов $C(n; 1, s)$ вычисляются по формулам $s = s_l^\pm = l(2d - 2 + i) \pm 1$, $1 \leq l < d/2$ и $(l, d) = 1$.

Д. Ф. Хсу и Дж. Шапиро [61] другим методом получили бесконечные семейства оптимальных величин n , аналогичные семействам из [109], и дали следующую верхнюю границу диаметра графа $C(n; 1, s)$:

Теорема 10. Для $n > 0$ справедливо неравенство $d_{\min}(n) < (n/2)^{\frac{1}{2}} + (n/8)^{\frac{1}{4}} + 2$.

В дополнение к этим классам величин n Дж.-К. Бермонд и Д. Цвиели [36], Э. А. Монахова [79], К. Мукхопадхья и Б. Синха [87] идентифицировали новые плотные бесконечные семейства оптимальных графов. Эти семейства покрывают все величины в $R(d)$ (без одной или двух), когда d или $d+1$ простое число [36, 79], и покрывают 92% всех величин n вплоть до $n = 8 \cdot 10^6$ [36]. Термин «плотный» используется здесь в том смысле, что семейства покрывают большие непрерывные интервалы в каждом $R(d)$, $d > 1$. Доказана

Теорема 11 [36]. Пусть $n \in R(d)$. Тогда n оптимально в следующих случаях:

- a) $(n, d) = 1$,
- b) $(n, d + 1) = 1$,
- c) $(n, d - 1) = 1$ и $n \leq 2d^2 + 1$.

Следствие 1. Если d или $d + 1$ простое число, то каждое n , такое, что $n \in R(d)$, $n \neq n_d - 1$, оптимально с возможным исключением $n = 2d^2 - 2$ в случае простого $d + 1$.

В [79] аналогичное свойство получено для предельно оптимальных графов.

Авторы [87], используя исчерпывающий поиск и аналог теоремы 11, показали, что предельно оптимальные описания могут быть получены более чем для 80% величин n , $n \leq 16000$, которые, следовательно, имеют гамильтонов цикл, связывающий все вершины графа.

Первый результат по аналитическому решению проблемы построения экстремальных двумерных кольцевых циркулянтов с заданным диаметром получен автором в [14]:

Теорема 12. Пусть $D \geq 1$ — целое число. Тогда $C(n_D; 1, 2D + 1)$ — предельно оптимальный циркулянт.

Этот результат был также получен Дж.-К. Бермондом с соавт. в [35] и Дж. Ибра с соавт. в [114], где показано, что $d(n_D; s_1, s_2) = D$, если и только если $s_1 \equiv Dp$ и $s_2 \equiv (D + 1)p$, где p — любое меньшее n и взаимно простое с n число. В последующие годы экстремальные графы из теоремы 12 были названы оптимальным семейством и активно исследовались [32, 39, 52, 77, 90, 91].

В 1990 г. Р. Браун и Р. Хогсон [39] получили оптимальное семейство $C(n_D; 1, 2D + 1)$, $D \geq 1$, нашли нижние границы диаметра и среднего расстояния и рассмотрели применение этой топологии в транзьютерной системе для обработки образов. В 2001 г. Л. Нараянан и др. [91], применяя семейство сетей $C(n_D; 1, 2D + 1)$, исследовали решение проблемы организации коммуникаций в оптических сетях, используя спектральное уплотнение каналов (или WDM-подход). Сеть в этом случае состоит из вершин, связанных волоконно-оптическими линиями связи, каждая из которых может поддерживать фиксированное число длин волн. В 2003 г. Р. Байвиде и др. [32, 77] исследовали семейство сетей $C(n_D; 1, 2D + 1)$ в качестве модели сети связи мультимодульных компьютерных систем. В работе [32] рассмотрены вложимость этих графов на поверхность

тора и возможность их будущего применения в системах с цилиндрическими структурами и атмосферными оптическими коммуникациями [69]; отмечено, что эти графы являются «идеальными» для применения диффузионных моделей коммуникаций, таких, как трансляционные и полные обмены. Авторы [32] рассмотрели также в качестве модели сетей связи семейство графов $C(2D^2; 1, 2D - 1)$, $D > 1$.

Недавно был сделан новый шаг в области изучения оптимальности графов $C(n; 1, s)$ [43, 44].

Пусть $n = qs + r$, $0 \leq r < s$, и $s = ar + b$, $0 \leq b < r$. Б. Чен и соавт. [44] получили формулы диаметра графов $C(n; 1, s)$ для подкласса значений n , а именно для $q > r$; для $q \leq r$ при $b \leq aq + 1$.

В 2006 г. Б. Чен и др. [43], используя эти формулы, нашли необходимые и достаточные условия того, что двумерный кольцевой циркулянтный граф порядка n , где $2d^2 + d < n < n_d - 1$, $d > 1$, имеет оптимальное описание.

Теорема 13. Предположим, что $n = 2t^2 + 2t - B$, где $t > B > 0$. Тогда существует положительное целое s , такое, что $d(n; 1, s) = t = D(n)$, если и только если существуют неотрицательные целые числа a, b, u и v , удовлетворяющие следующим условиям:

- 1) $n = av + bu$;
- 2) $a + b \leq 2t + 1$ и $u + v \leq 2t + 1$;
- 3) $b \geq a$, $u \geq v$, $u > a$, $v < b$;
- 4) справедливо одно из следующих двух условий:
 - а) $x = 0$, $b - a = y$ — нечетное число, $y | 2B + 1$ и $1 \leq y \leq \sqrt{2B + 1}$,
 - б) $y = 0$, $u - v = x$ — нечетное число, $x | 2B + 1$ и $1 \leq x \leq \sqrt{2B + 1}$, где $x = 2t + 1 - (a + b)$, $y = 2t + 1 - (u + v)$;
- 5) $(b, v) = 1$.

Применяя эти условия, авторы получили ряд новых оптимальных и субоптимальных бесконечных семейств двумерных кольцевых циркулянтных графов. Но проблема классификации всех значений n , для которых такие графы существуют, остается открытой. Перечислим некоторые из задач, которые еще не решены.

Задача 4. Получение формулы для диаметра двумерной кольцевой сети в случае $q \leq r$, когда $b > aq + 1$ [43].

Задача 5. Улучшение оценок, найденных для $d(n, k)$.

Задача 6. Получение вида функции $d(n, k)$, когда значение k фиксировано.

Известно, что когда значение k фиксировано, $d(n, k)$ не есть монотонно увеличивающаяся функция от n . Например, для $k = 3$ имеем $d(103, 3) = 5$, но $d(117, 3) = 4$. В работе [41] показано компьютерным поиском, что $d(n, 3) \leq d(n', 3) + 1$ для всех $n < n' \leq 1000$. В [109] проверено, что $d(n, 2) \leq d(n', 2) + 1$ для всех $n < n' \leq 8 \cdot 10^6$.

5. Трехмерные циркулянты

В случае размерности $k = 3$ из [111] следует

$$(8/27)d^3 + O(d^2) \leq M(d, 3) \leq 1 + (4d^3 + 6d^2 + 8d)/3.$$

Дж. Ибра и др. [114] доказали для $k = 3$, что максимальный порядок циркулянта $C(n; s_1, s_2, s_1 + s_2)$ диаметра d , где $1 \leq s_1 < s_2 < \lfloor n/2 \rfloor$, равен $3d^2 + 3d + 1$. Данная граница достигается для семейства графов с $s_1 = 1$ и $s_2 = 3d + 1$. Это семейство изучали К. Шин [104] и Ф. Агуило с соавт. [27]. В 2008 г. К. Мартинес и др. [11] применили семейство из [114] к конструированию совершенных кодов, корректирующих ошибки.

Отметим, что результаты, полученные К. Гарсиа и П. Соле в [54], показывают, что для увеличения порядка графа n необходимо использовать образующие, линейно независимые над \mathbb{Z}_n .

К. Пейра (см. заметку в [33]) нашла семейство трехмерных кольцевых циркулянтов диаметра d с порядком $n = (8/9)d^3 + O(d^2)$. Ш. Чен и Х.-Д. Джиа [41] и К. Делорме (см. [33]) нашли семейство трехмерных кольцевых циркулянтов порядка n с диаметром, не превосходящим $d \geq 3$, и $M(d, 3) \geq n = 32[d/3]^3 + 8[d/3]^2 + 2[d/3]$.

В 2003 г. получен следующий результат [81]:

Теорема 14. Максимальный порядок трехмерного кольцевого циркулянта $C(n; 1, s_2, s_3)$ диаметра $d \geq 1$ есть

$$M(d, 3) = \begin{cases} \frac{32}{27}d^3 + \frac{16}{9}d^2 + 2d + 1, & \text{если } d \equiv 0 \pmod{3}, \\ 32[d/3]^3 + 48[d/3]^2 + 30[d/3] + 7, & \text{если } d \equiv 1 \pmod{3}, \\ 32[d/3]^3 + 80[d/3]^2 + 70[d/3] + 21, & \text{если } d \equiv 2 \pmod{3}, \end{cases}$$

и достигается при следующих образующих:

$$(s_2, s_3) = \begin{cases} (\frac{8}{9}d^2 + \frac{2}{3}d, \frac{8}{9}d^2 + 2d + 2), & \text{если } d \equiv 0 \pmod{3}, \\ (8[d/3]^2 + 6[d/3] + 2, 8[d/3]^2 + 10[d/3] + 4), & \text{если } d \equiv 1 \pmod{3}, \\ (8[d/3]^2 + 10[d/3] + 4, 8[d/3]^2 + 14[d/3] + 6), & \text{если } d \equiv 2 \pmod{3}. \end{cases}$$

Для описания построенных графов найдено три различных множества образующих (по крайней мере, для $d \equiv 0 \pmod{3}$ и $d \equiv 2 \pmod{3}$) [18, 81]. В табл. 2 представлены описания трехмерных кольцевых циркулянтных графов с максимальным порядком, равным $M(d, 3)$, для диаметров $1 \leq d \leq 18$.

Таблица 2

Трехмерные кольцевые циркулянты с максимальным порядком

d	$M(d, 3)$	s_2, s_3	s_2, s_3	s_2, s_3	d	$M(d, 3)$	s_2, s_3	s_2, s_3	s_2, s_3
1	7	2, 4			10	1393	92, 106		
2	21	4, 6	3, 8	5, 9	11	1815	106, 120	15, 241	839, 855
3	55	10, 16	5, 21	20, 24	12	2329	136, 154	17, 273	1088, 110
4	117	16, 22			13	2943	154, 172		
5	203	22, 28	7, 57	83, 91	14	3629	172, 190	19, 381	1709, 1729
6	333	36, 46	9, 73	144, 152	15	4431	210, 232	21, 421	2100, 2120
7	515	46, 56			16	5357	232, 254		
8	737	56, 66	11, 133	329, 341	17	6371	254, 276	23, 553	3035, 3059
9	1027	78, 92	13, 157	468, 480	18	7525	300, 326	25, 601	3600, 3624

С помощью компьютерного поиска для $k = 3$ проверено, что при $n > M(d, 3)$, $2 \leq d \leq 6$, циркулянты диаметра d не существуют. Таким образом, можно высказать следующую гипотезу: максимально достижимый порядок трехмерного циркулянта $C(n; s_1, s_2, s_3)$ диаметра $d \geq 1$, где $1 \leq s_1 < s_2 < s_3 < n$, равен $M(d, 3)$.

6. Мультипликативные циркулянтные графы

Во многих работах исследуется диаметр многомерных циркулянтов, а также максимально достижимый их порядок как функция от d и k . Первый результат по изучению многомерных циркулянтов был получен в 1974 г.

Теорема 15 [111]. Для нечетного $s > 1$ диаметр циркулянта $C(s^k; 1, s, s^2, \dots, s^{k-1})$ равен $\frac{k}{2}n^{1/k} - \frac{k}{2} = k\lfloor s/2 \rfloor$; среднее расстояние равно $k(s^2 - 1)/4s$.

Ш. Чен и Х.-Д. Джиа в 1993 г. доказали следующий важный результат:

Теорема 16 [41]. Пусть d и k — любые положительные целые, такие, что $d \geq k \geq 3$, и пусть $q = \lfloor (d - k + 3)/k \rfloor$. Тогда

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

При доказательстве теоремы 16 авторы определили, что верхняя граница диаметра циркулянтов вида $C(M; 1, 4q, \dots, (4q)^{k-1})$ равна $k(q + 1) - 3$.

Будем называть, следуя [105], *мультипликативными* циркулянтные сети вида $C(n; 1, s, s^2, \dots, s^{k-1})$ с образующими в виде степеней некоторого числа $s \geq 2$. В этом определении мультипликативных циркулянтов величина n может быть произвольной, не обязательно равной s^k . Заметим, что семейство из [111] — мультипликативные циркулянты с нечетным s , а из [41] — с четным s , кратным 4.

И. Стойменович (1997 г.) [105] изучил топологические свойства мультипликативных циркулянтов с четным s и порядком $n = s^k$ и получил их диаметр и среднее расстояние.

Теорема 17 [105]. Диаметр циркулянтной сети $C(s^k; 1, s, s^2, \dots, s^{k-1})$, где $s > 2$ — четное число, равен $ks/2 - \lfloor k/2 \rfloor$.

Б. Пархами в 2006 г. [93, 94] описал структурные свойства мультипликативных циркулянтных сетей с нечетным $s \geq 3$ и порядками $2s^{k-1} < n \leq s^k$ и определил формулу для их диаметра.

Теорема 18 [93, теорема 6]. Диаметр циркулянтной сети $C(n; 1, s, s^2, \dots, s^{k-1})$, где $s > 1$ — нечетное число и $n > 2s^{k-1}$, равен

$$(k - 1)\lfloor s/2 \rfloor + \lceil (n - s^{k-1})/(2s^{k-1}) \rceil. \quad (3)$$

И. Стойменович и Б. Пархами также показали, что мультипликативные циркулянтные сети имеют простые оптимальные алгоритмы парного [93, 94, 105] и трансляционного обменов [105].

Другие структурные характеристики, относящиеся к диаметру, получены в [71] для ориентированных мультипликативных циркулянтов вида $G(s^k; 1, s, \dots, s^{k-1})$. К. Гарсия и П. Соле [54] дали более точные оценки диаметра, чем в [111], для графов вида $C(n; 1, s, \dots, s^{k-1})$, когда n — простое число. Ф. Хванг [65] рассмотрел проблему сокращения диаметра по сравнению с [111] для ориентированных мультипликативных циркулянтов любого порядка n .

Диаметры графов семейств мультипликативных циркулянтов размерностей $k \geq 4$ исследованы автором в работах [19, 20]. Построены бесконечные семейства циркулянтов, достигающих найденных границ диаметра.

Теорема 19 [19]. Пусть $k > 2$ — целое, $s > 2$ — нечетное число и $S = \{1, s, s^2, \dots, s^{k-1}\}$. Если $n = \lceil s/2 \rceil \sum_{i=0}^{k-1} s^i$, то $d(n; S) = \left\lfloor \frac{k}{2} \lceil s/2 \rceil \right\rfloor$.

Теорема 20 [19]. Пусть $k > 2$ — целое, $s > 2$ — нечетное число и $S = \{1, s, s^2, \dots, s^{k-1}\}$. Если $n = \lceil s/2 \rceil \sum_{i=0}^{k-2} s^i + \lfloor s/2 \rfloor s^{k-1}$, то

$$d(n; S) = \begin{cases} \frac{k}{2} \lceil s/2 \rceil - 1 & \text{для четного } k \text{ или } s \equiv 3 \pmod{4}, \\ \left\lfloor \frac{k}{2} \lceil s/2 \rceil \right\rfloor & \text{в остальных случаях.} \end{cases}$$

Применяя теорему 20 для $k = 4$, получим семейство циркулянтов, которые улучшают известные оценки экстремальной функции $M(d, 4)$.

Теорема 21. Пусть $d > 4$ — целое и $q = 2\lfloor (d-1)/2 \rfloor + 1$. Тогда

$$M(d, 4) \geq n = \frac{q^4 + 1}{2} + q^2 + q.$$

Сравнение формулы (3) с диаметрами графов семейств из теорем 19 и 20 показывает, что (3) справедлива при $s = 3$ и $k \geq 3$ для семейства из теоремы 19. Тем не менее формула (3) не имеет места для найденных семейств графов при $k \geq 4$ и нечетном $s > 3$ и, следовательно, дает только верхнюю границу диаметра.

В работе [20] улучшена оценка диаметра графов семейства из [41]. Вместе с результатом теоремы 19 это дает новые нижние границы экстремальной функции $M(d, k)$ для всех $k > 4$.

Теорема 22 [20]. Пусть $p = \lfloor (d - \lfloor k/4 \rfloor)/k \rfloor$, где d и $k > 4$ — целые числа, такие, что $d \geq k + \lfloor k/4 \rfloor$. Тогда

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

Пример 2. Пусть $k = 7$. Тогда в силу теоремы 16

$$M(d, 7) \geq 10922 \text{ для } 11 \leq d \leq 17.$$

Теорема 22 дает следующие оценки:

$$M(d, 7) \geq n = \begin{cases} 58593 & \text{для } 10 \leq d \leq 13, \\ 549028 & \text{для } d = 14, \\ 1198372 & \text{для } 15 \leq d \leq 16, \\ 2989355 & \text{для } d = 17. \end{cases}$$

Указанные значения n и нижние границы d достигаются для образующих $S = \{1, s, s^2, s^3, s^4, s^5, s^6\}$ со значениями s , равными 5, 7, 8 и 9 соответственно.

Для дальнейших исследований предлагаем следующие открытые вопросы.

Задача 7. Улучшить найденные выше оценки достижимого порядка циркулянтных графов размерностей $k \geq 4$ и найти семейства графов, достигающих эти оценки.

Задача 8. Разработать аналитические методы решения задачи поиска кратчайших путей для таких сетей с минимальной структурной задержкой.

7. Методы синтеза оптимальных циркулянтов

Основными методами, используемыми для построения регулярных графов с минимальным диаметром и/или минимальным средним расстоянием, являются локальный поиск и поиск в ширину: работы М. Имаса и М. Итона [66], В. В. Корнеева [9], О. Г. Монахова [12], Х. Стоуна [106], С. Тога и К. Стейглица [107], Ш. Чена и Х.-Д. Джиа [41] (в последнем случае — синтез трехмерных циркулянтов). Ограничением этих алгоритмов является большое время исполнения при больших значениях числа вершин n и размерности k .

Для произвольных циркулянтных графов с $k \geq 3$ трудной проблемой является определение существования оптимального графа для заданных n и k . В работе [41] найдены интервалы существования оптимальных графов для $k = 3$ и $n \leq 1000$ и показано, что по крайней мере субоптимальные графы существуют для всех рассмотренных n . В [21] аналогичные результаты получены для $k = 3$ и $n \leq 10000$. Что касается произвольных значений $k \geq 3$, то бесконечные семейства оптимальных многомерных циркулянтов найдены только для диаметра, равного двум [17].

В литературе описано также несколько эвристических алгоритмов синтеза циркулянтов с минимальным диаметром и/или минимальным средним расстоянием. С. В. Труфанов [24] получил алгоритм для сокращения диаметра максимально связанных графов, включая циркулянтные графы. В. А. Воробьев [3] предложил алгоритм синтеза для двумерных циркулянтов. Д. Цвиели [109] получила границы оптимальных образующих для двумерных циркулянтов $C(n; 1, s)$ и алгоритм их вычисления, если они существуют. Автором разработан алгоритм синтеза оптимальных (предельно оптимальных) циркулянтов размерностей $k \geq 2$ и получен каталог этих сетей для $2 \leq k \leq 5$ [14].

Новый подход, основанный на задании темплейтов и множества входных-выходных пар и использующий эволюционные вычисления, предложен О. Г. Монаховым [13] и использован в [22, 82, 83] для синтеза графов. Представленный алгоритм эволюционного синтеза интегрировал преимущества генетических алгоритмов и генетического программирования и был применен для автоматического переоткрытия и открытия некоторых графовых, вычислительных и комбинаторных алгоритмов. Основная идея алгоритма состоит в эволюционных преобразованиях над множествами аналитических описаний графов (формул), основанных на естественной селекции: выживает «сильнейший». Функция пригодности оценивает в одном случае семейство графов с максимальным числом вершин для заданных диаметров, в другом — сумму диаметров для циркулянтных графов с заданными размерностью, множествами образующих и порядками.

Алгоритм переоткрыл известные семейства графов. Кроме того, для семейства из [41] был найден новый вид образующих. Для $k = 3$ и $k = 4$ алгоритм автоматически породил описания новых неизвестных ранее семейств, которые соответствуют циркулянтным графам с лучшими экстремальными свойствами для большой области изменения диаметров.

8. Парные обмены в циркулянтах

Большинство семейств циркулянтов, описанных в литературе, изучаются как сети связи мультипроцессорных систем. Их коммуникационные свойства также очень важны.

При парной маршрутизации сообщение посылается из вершины-источника в вершину-приемник. Это одна из наиболее важных коммуникационных проблем для лю-

бой сети. Алгоритм парной маршрутизации оптимален, если сообщение посылается между любыми двумя вершинами вдоль кратчайшего пути. Есть два типа стратегий парной маршрутизации: коммутация сообщений (store-and-forward routing) и коммутация каналов (wormhole routing). При *коммутации сообщений* сообщение полностью находится в вершине перед передачей в следующую вершину на пути к приемнику. При *коммутации каналов* сначала в сети выделяется весь путь из источника в приемник посредством передачи заголовка сообщения, затем сообщение передается вдоль настроенных каналов между источником и приемником. Один и тот же алгоритм парной маршрутизации может быть адаптирован для обеих стратегий.

Известен алгоритм Дейкстры поиска кратчайшего пути с квадратичной сложностью $O(n^2)$, применимый для любого связного графа порядка n . Дж. Кей с соавт. [40] установили, что в произвольных циркулянтных графах проблема поиска кратчайшего пути с использованием множества образующих $S = \{s_1, s_2, \dots, s_k\}$ является NP-трудной. Предложены различные алгоритмы поиска кратчайших путей с допустимыми оценками для двумерных циркулянтных сетей [16, 42, 52, 57, 87, 90, 99, 100], а также для некоторых семейств трехмерных циркулянтов.

Для размерности $k = 2$ К. Мукхопадхья и Б. Синха (1995 г.) [87] дали отказоустойчивый алгоритм в сетях вида $C(n; 1, s)$ с оценкой $O(d)$, где d — диаметр графа, и субоптимальный алгоритм — превышающий оптимальный не более чем на единицу в случае единичного отказа вершины или ребра. Для графов вида $C(n; s_1, s_2)$, $1 \leq s_1 < s_2 < n$, Б. Робич (1996 г.) [99] разработал алгоритм парных обменов по кратчайшим путям. Алгоритм требует $O(d)$ времени для инициализации и $l = O(d)$ шагов, где l — расстояние между вершинами, а d — диаметр сети. Д. Гомес и др. (2007 г.) [57] представил полиномиальной сложности алгоритм вычисления кратчайшего пути между двумя вершинами циркулянтного графа степени четыре. Он требует $O(\log^3 n)$ бит операций в графе порядка n и основан на определении вектора кратчайших путей в специальном классе решеток для L_1 -норм. В 2005 г. Б. Чен с соавт. [42] представили оптимальный алгоритм для двумерных циркулянтов с константной сложностью, требующий предопределения четырех параметров укладки в виде буквы L графа $C(n; s_1, s_2)$ (L -shape tile [111]) и решения сравнения $s_1x + s_2y \equiv 1 \pmod{n}$.

Для оптимальных графов $C(n; D, D + 1)$, $D \geq 1$, любых порядков в [16] получен отказоустойчивый алгоритм поиска кратчайших путей со сложностью $O(1)$. В 1997 г. Й. Фабрега и М. Зарагоса [52] исследовали оптимальные графы $C(n_D; 1, 2D + 1)$ и $C(n_D; D, D + 1)$ и получили покрывающее дерево кратчайших путей и оценки отказоустойчивости предложенного алгоритма маршрутизации. К. Мартинес и др. [78], используя пометку вершин тороидальных графов посредством гауссианских целых, получили, в частности, алгоритм поиска кратчайшего пути в циркулянтах вида (1) с $n = n_D$ с оценкой $O(D)$. Эти авторы в 2007 г. показали, что расстояние в двумерных циркулянтах является подходящей метрикой для того, чтобы конструировать совершенные коды, исправляющие ошибки. В 2009 г. Б. Б. Нестеренко и М. А. Новотарский [23] разработали алгоритм оптимального парного обмена в клеточных нейронных сетях для семейства циркулянтов вида $C(n; \sqrt{n}/2 - 1, \sqrt{n})$, $n = 2^{2i}$, $i \geq 3$. Алгоритм включает прямые и альтернативные маршруты между двумя нейронами сети.

Следует отметить, что вычисление кратчайших путей в оптимальных циркулянтных сетях степени четыре требует меньшего числа операций, чем аналогичная процедура, найденная Б. Арденом и Х. Ли в [28] для хордальных кольцевых сетей степени три, которые являются подграфами циркулянтов.

В работах [73, 90] изучается возможность эффективного использования интервальной маршрутизации в циркулянтных сетях. Интервальная маршрутизация основана на представлении таблиц маршрутизации, находящихся в каждой вершине, в компактной форме. В 1997 г. Л. Нараянан и Я. Опатни [90] исследовали схемы компактной и интервальной маршрутизации в графах $C(n; 1, s)$, включая графы $C(n_D; 1, 2D + 1)$. Схема маршрутизации для любой такой сети требует $O(\log n)$ бит информации в каждой вершине и $O(1)$ времени для вычисления кратчайшего пути до любого приемника. Даны границы для схем интервальной маршрутизации таких сетей. В 1999 г. Б. Манс [73] использовал результаты из [90] для рассмотрения интервальной маршрутизации в графах общего вида $C(n; 1, s)$. Точное определение интервальной маршрутизации и более детальную информацию о ней можно найти в [55].

Для размерности $k = 3$ А. Листман с соавт. [72] исследовали сетевые свойства трехмерных циркулянтов, используя в качестве модели сети семейства циркулянтов с $n = O(3d^2)$ из [114], и получили решение проблемы поиска кратчайших путей по описанию графа. Л. Барриере с соавт. [30] нашли алгоритм поиска кратчайших путей в этих сетях. В [72] решена проблема кратчайших путей для эквивалентного графа вида $C(3d^2 + 3d + 1; d, d + 1, 2d + 1)$. Автором показано [81], что проблема поиска кратчайших путей для семейства трехмерных циркулянтов максимального порядка при заданном диаметре имеет сложность $O(1)$ и решается простым аналитическим методом. Получена также функция расстояний для этих графов.

9. Трансляционные и полные обмены

Организация трансляционных и полных обменов является одной из наиболее актуальных задач в сетевых коммуникациях и параллельных вычислительных системах (см., например, [10, 21, 59, 96]). При *трансляционном* обмене (broadcasting) сообщение, первоначально содержащееся в одной вершине сети (называемой источником), должно быть передано во все другие вершины. При *полном* обмене (gossiping) каждая вершина содержит сообщение, которое должно быть передано во все другие вершины. Эти типы сетевых коммуникаций имеют место при решении многих проблем параллельных вычислений, а также при глобальной процессорной синхронизации или обновлении распределенных баз данных. Поэтому большое количество работ посвящено поиску эффективных алгоритмов этих коллективных обменов. Основными мерами эффективности данных алгоритмов являются число требуемых шагов (время) и число элементарных передач между парами вершин.

Трансляционные обмены в оптимальных двумерных циркулянтах исследовались в [16, 45, 72, 78, 80, 92] при различных коммуникационных моделях. В 1982 г. автором представлен алгоритм трансляционного обмена для оптимальных графов вида (1) любого порядка с временем, равным диаметру графа [16]. Результат получен в n -портовой модели (shouting model), когда вершина может информировать всех своих соседей в течение одной единицы времени. Важной особенностью алгоритма является то, что он работает без дублирования пакетов и использует для каждой транзитной вершины i покрывающее дерево кратчайших путей с центром в i . К. Мартинес и др. [78], используя алгебраический подход, получили в n -портовой модели алгоритм трансляционного обмена для всех гауссианских сетей, включая циркулянты вида (1) с $n = n_D$ с оценкой D , где D — диаметр сети. В 1998 г. А. Листман и др. [72] определили, что время трансляционного обмена для этих сетей равно $D + 2$. Эти результаты получены в 1-портовой модели (whispering model): в течение каждой единицы времени вершина может информировать не более одного соседа.

Рассматривая варианты трансляционных обменов, которые имеют сходство с распространением компьютерных вирусов в сетях, Ф. Комеллас и соавт. [45] в 2002 г. получили конструкции оптимальных деревьев трансляционного обмена и оценку его времени для оптимальных двумерных циркулянтов из [38] в коммуникационной модели, когда каждая вершина может информировать i , $2 \leq i \leq 4$, вершин во время каждого шага, когда она активна. В 2005 г. Н. Обрадович с соавт. [92] ввели пересмотренное по сравнению с [72] и основанное на отказоустойчивости определение оптимальности дерева трансляционного обмена для i -портовой модели. Они дали конструкции оптимальных деревьев трансляционного обмена для i -портовых сетей, $1 \leq i \leq 4$, моделируемых оптимальными графами $C(n; D, D+1)$ диаметра D , где $n \in R(D)$.

Для размерности $k = 3$ в [72] получено время трансляционного обмена, равное $d + 3$ для циркулянтов вида $C(3d^2 + 3d + 1; d, d + 1, 2d + 1)$ диаметра d в 1-портовой модели. Трехмерные графы из теоремы 14 с максимально возможным порядком для заданного диаметра изучались как коммуникационная сеть в работах [18, 58]. Автором представлен [18] алгоритм трансляционного обмена для этих сетей с временем, равным диаметру d графа в n -портовой модели. Х. Харутюнян и Э. Марачлян [58] в 2008 г. получили алгоритмы и оценки трансляционного обмена в этих сетях в 1-портовой модели и доказали, что $d + 2$ есть нижняя граница времени обмена. Результаты получены с использованием следующей теоремы.

Теорема 23. Если циркулянт C диаметра d имеет более чем $d + 2$ вершины на расстоянии d от вершины 0, то время трансляционного обмена в C удовлетворяет неравенству $b(C) \geq d + 2$.

Для многомерных циркулянтных графов любой степени $2k$ они также показали, что $d + 2k - 1$ есть верхняя граница времени трансляционного обмена, и представили алгоритм, который информирует все вершины любого циркулянта степени $2k$ и диаметра d самое большее за $d + 2k - 1$ единиц времени.

Для произвольной размерности k Ф. Комеллас и др. [45] рассмотрели приложение циркулянтных сетей в модели «малого мира» и получили оценки времени трансляционного обмена для циркулянтов степени $v = 2k$ и вида $C(n; 1, 2, \dots, k)$ в коммуникационной модели, когда каждая вершина может информировать $i \leq v$ вершин во время каждого шага, когда она активна.

Полный обмен в циркулянтных сетях изучался автором [16, 80] для предельно оптимальных двумерных циркулянтов. Б. Манс и И. Шпарлинский [76] предложили при решении проблемы организации полных обменов в циркулянтных графах использовать знания о ширине разреза этих графов.

10. Рекурсивные циркулянты

Коротко опишем графы, относящиеся к рекурсивным циркулянтам. В 1994 г. Ю. Парк и К. Чва [95] предложили новую топологию для мультикомпьютерных сетей, названную *рекурсивными циркулянтами*.

Определение 8. Рекурсивный циркулянт $G(N, s)$ есть циркулянтный граф вида $C(N; 1, s, s^2, \dots, s^{\lceil \log_s N \rceil - 1})$, $s \geq 2$.

В [95] рассмотрено подмножество класса рекурсивных циркулянтов с порядком $N = cs^m$, где c и s — целые числа и $1 \leq c < s$. Для любого $s \geq 3$ рекурсивные циркулянты $G(cs^m, s)$ являются регулярными графами степеней $2m$, если $c = 1$; $2m + 2$, если $c > 2$; $2m + 1$, если $c = 2$. Для введенного класса графов авторы в [95] определили диаметр, среднее расстояние, рассмотрели связность и предложили простой алгоритм

поиска кратчайшего пути. Позднее ряд авторов (A. Raspaud, G. Fertin, C. Micheneau, D. Biss и др.) изучили различные свойства рекурсивных циркулянтов, включая распознавание, вложимость, свойство гамильтоновости и организацию трансляционных обменов в них.

Интересно отметить, что при $c = 1$ рекурсивные циркулянты являются мультипликативными циркулянтными графами, которые рассматривались в работах [93, 94, 105, 111].

Автор благодарен всем своим коллегам за плодотворное сотрудничество, а также О. Г. Монахову за полезное обсуждение настоящей работы.

ЛИТЕРАТУРА

1. Артамонов Г. Т. Топология регулярных вычислительных сетей и сред. М.: Радио и связь, 1985. 192 с.
2. Артамонов Г. Т., Тюрин В. Д. Топология сетей ЭВМ и многопроцессорных систем. М.: Радио и связь, 1991. 248 с.
3. Воробьев В. А. Простейшие структуры однородных вычислительных систем // Вычислительные системы. Вопросы теории и построения ВС. Новосибирск, 1974. № 60. С. 35–49.
4. Воробьев В. А., Корнеев В. В. Некоторые вопросы теории структур однородных вычислительных систем // Вычислительные системы. Вопросы теории и построения ВС. Новосибирск, 1974. № 60. С. 3–16.
5. Димитриев Ю. К. Анализ самодиагностических свойств структур распределенных живучих вычислительных систем // Автометрия. 1996. № 5. С. 71–84.
6. Ебдокимов С. А., Пономаренко И. Н. Распознавание и проверка изоморфизма циркулянтных графов за полиномиальное время // Алгебра и анализ. 2003. Т. 15. № 6. С. 1–34.
7. Евреинов Э. В., Хорошевский В. Г. Однородные вычислительные системы. Новосибирск: Наука, 1978. 318 с.
8. Клейнрок Л. Коммуникационные сети. Стохастические потоки и задержки сообщений. М.: Наука, 1970. 256 с.
9. Корнеев В. В. О макроструктуре однородных вычислительных систем // Вычислительные системы. Вопросы теории и построения ВС. Новосибирск, 1974. № 60. С. 17–34.
10. Корнеев В. В. Параллельные вычислительные системы. М.: Нолидж, 1999. 320 с.
11. Мартинес К., Стаффорд Э., Байвиде Р., Габидулин Э. М. Представление гексагональных созвездий с помощью графов Эйзенштейна — Якоби // Проблемы передачи информации. 2008. Т. 44. № 1. С. 3–13.
12. Монахов О. Г. Параметрическое описание структур однородных вычислительных систем // Вычислительные системы. Вопросы теории и построения ВС. Новосибирск, 1979. № 80. С. 3–17.
13. Монахов О. Г. Эволюционный синтез алгоритмов на основе шаблонов // Автометрия. Новосибирск, 2006. Т. 42. № 1. С. 116–126.
14. Монахова Э. А. Синтез оптимальных диофантовых структур // Вычислительные системы. Вопросы теории и построения ВС. Новосибирск, 1979. № 80. С. 18–35.
15. Монахова Э. А. Об аналитическом описании оптимальных двумерных диофантовых структур однородных вычислительных систем // Вычислительные системы. Вопросы теории и построения ВС. Новосибирск, 1981. № 90. С. 81–91.

16. Монахова Э. А. Алгоритмы межмашинных взаимодействий и реконфигурации графов связей в вычислительных системах с программируемой структурой // Вычислительные системы. Вопросы теории и построения ВС. Новосибирск, 1982. № 94. С. 81–102.
17. Монахова Э. А. Оптимальные КАИС-структуры однородных вычислительных систем // Электронное моделирование. 1985. № 3. С. 30–34.
18. Монахова Э. А. Трехмерные циркулянтные сети связи параллельных вычислительных систем // Автометрия. 2006. № 3. С. 106–118.
19. Монахова Э. А. Мультипликативные циркулянтные сети // Дискрет. анализ и исслед. опер. 2010. Т. 17. № 5. С. 56–66.
20. Монахова Э. А. Об одном экстремальном семействе циркулянтных сетей // Дискрет. анализ и исслед. опер. 2011. Т. 18. № 1. С. 66–76.
21. Монахов О. Г., Монахова Э. А. Параллельные системы с распределенной памятью: структуры и организация взаимодействий. Новосибирск: Изд-во СО РАН, 2000. 242 с.
22. Монахов О. Г., Монахова Э. А. Синтез новых семейств оптимальных регулярных сетей на основе эволюционных вычислений и темплейтов функций // Автометрия. 2004. № 4. С. 106–116.
23. Нестеренко Б. Б., Новотарский М. А. Клеточные нейронные сети на циркулянтных графах // Искусственный интеллект. 2009. № 3. С. 132–138.
24. Труфанов С. В. Некоторые задачи о расстояниях на графе // Изв. АН СССР. Техн. кибернетика. 1967. № 3. С. 61–66.
25. Яблонский С. В. Алгоритм построения вычислительных сетей с минимальным средним расстоянием между узлами // Тез. докл. Всес. совещ. «Методы и программы решения оптимизационных задач на графах и сетях». Новосибирск, 1980. С. 103–105.
26. *Ádám A.* Research problem 2–10 // J. Combin. Theory. 1967. No. 2. P. 393.
27. *Aguilo F., Fiol M. A., and Garcia C.* Triple Loop Networks with Small Transmission Delay // Discrete Math. 1997. V. 167/168. P. 3–16.
28. *Arden B. W. and Lee H.* Analysis of chordal ring networks // IEEE Trans. Computers. 1981. No. C-30. P. 291–295.
29. *Balaban A. T.* Reaction graphs // Graph Theoretical Approaches to Chemical Reactivity / eds. D. Bonchev and O. Mekenyan. Netherlands: Kluwer Academic Publishers, 1994. P. 137–180.
30. *Barriere L., Fabrega J., Simo E., and Zaragoza M.* Fault-Tolerant Routings in Chordal Ring Networks // Networks. 2000. V. 36. No. 3. P. 180–190.
31. *Beivide R., Herrada E., Balcazar J. L., and Arruabarrena A.* Optimal distance networks of low degree for parallel computers // IEEE Trans. Computers. 1991. V. 40. No. 10. P. 1109–1124.
32. *Beivide R., Martinez C., Izu C., et al.* Chordal Topologies for Interconnection Networks // LNCS. 2003. V. 2858. P. 385–392.
33. *Bermond J.-C., Comellas F., and Hsu D. F.* Distributed loop computer networks: a survey // J. Parallel Distributed Comput. 1995. V. 24. P. 2–10.
34. *Bermond J.-C., Favaron O., and Maheo M.* Hamiltonian decomposition of Cayley graphs of degree four // J. Combin. Theory. Ser. B. 1989. No. 46. P. 142–153.
35. *Bermond J.-C., Iliades G., and Peyrat C.* An optimization problem in distributed loop computer networks // Third Inter. Conf. Combinatorial Math. New York, USA, June 1985. Ann. New York Acad. Sci., 1989. No. 555. P. 45–55.
36. *Bermond J.-C. and Tzvieli D.* Minimal diameter double-loop networks: Dense optimal families // Networks. 1991. No. 21. P. 1–9.

37. *Boesch F. T. and Tindell R.* Circulants and their connectivity // J. Graph Theory. 1984. No. 8. P. 487–499.
38. *Boesch F. T. and Wang J.-F.* Reliable circulant networks with minimum transmission delay // IEEE Trans. Circuits Syst. 1985. No. CAS-32. P. 1286–1291.
39. *Browne R. F. and Hodgson R. M.* Symmetric degree-four chordal ring networks // IEE Proc. 1990. V. 137. No. 4. P. 310–318.
40. *Cai J.-Y., Havas G., Mans B., et al.* On Routing in Circulant Graphs // LNCS. 1999. V. 1627. P. 360–369.
41. *Chen S. and Jia X.-D.* Undirected loop networks // Networks. 1993. No. 23. P. 257–260.
42. *Chen B.-X., Meng J.-X., and Xiao W.-J.* A Constant Time Optimal Routing Algorithm for Undirected Double-Loop Networks // First Int. Conf. Mobile Ad-hoc and Sensor Networks. MSN 2005, Wuhan, China, December 2005. P. 309–316.
43. *Chen B.-X., Meng J.-X., and Xiao W.-J.* Some new optimal and suboptimal infinite families of undirected double-loop networks // DMTCS. 2006. No. 8. P. 299–312.
44. *Chen B.-X., Xiao W.-J., and Parhami B.* Diameter formulas for a class of undirected double-loop networks // J. Intercon. Networks. 2005. V. 6. No. 1. P. 1–15.
45. *Comellas F., Mitjana M., and Peters J. G.* Broadcasting in Small-World Communication Networks // 9th Inter. Coll. on Structural Information and Communication Complexity (SIROCCO 9), 2002. Proc. Informatics. 2002. No. 13. P. 73–85.
46. *David H. A.* Enumeration of cyclic graphs and cyclic designs // J. Comb. Theory. 1972. No. 13. P. 303–308.
47. *Davis P. J.* Circulant Matrices. New York: Wiley Publ., 1979. 304 p.
48. *Delorme C. and Maheo M.* Isomorphisms of cayley multigraphs of degree four on finite abelian groups // Eur. J. Combinat. 1992. No. 13. P. 59–61.
49. *Du D.-Z., Hsu D. F., Li Q., and Xu J.* A combinatorial problem related to distributed loop networks // Networks. 1990. No. 20. P. 173–180.
50. *Elspas B.* Topological constructions on interconnection limited logic // Switch. Circ. Theor. Log. Des. 1964. No. 164. P. 133–147.
51. *Elspas B. and Turner J.* Graphs with circulant adjacency matrices // J. Comb. Theory. 1970. No. 9. P. 229–240.
52. *Fabrega J. and Zaragoza M.* Fault-tolerant routings in double fixed-step networks // Discr. Appl. Math. 1997. No. 78. P. 61–74.
53. *Feng X. and Xu M.* On isomorphisms of Cayley graphs of small valency // Algebra Colloquium. 1994. V. 1. P. 67–76.
54. *Garcia C. and Solé P.* Diameter lower bound for Waring graphs and multiloop networks // Discr. Math. 1993. No. 111. P. 257–261.
55. *Gavoille C.* A survey on interval routing // Theor. Comp. Sci. 2000. No. 245. P. 217–253.
56. *Gobel F. and Neutel E. A.* Cyclic graphs // Discr. Appl. Math. 2000. No. 99. P. 3–12.
57. *Gomez D., Gutierrez J., Ibeas A., and Beivide R.* Optimal routing in double loop networks // Theor. Comp. Sci. 2007. V. 381. Issue 1–3. P. 68–85.
58. *Harutyunyan H. A. and Maraachlian E.* Near Optimal Broadcasting in Optimal Triple Loop Graphs // IEEE 22nd Inter. Conf. on Advanced Information Networking and Applications, AINA 2008, March 25–29, Ginowan, Okinawa, Japan. P. 167–181.
59. *Hedetniemi S. M., Hedetniemi S. T., and Liestman A. L.* A survey of gossiping and broadcasting in communication networks // Networks. 1988. No. 18. P. 319–349.
60. *Hsu D. F. and Jia X. D.* Extremal problems in the combinatorial construction of distributed loop networks // SIAM J. Discr. Math. 1994. No. 7. P. 57–71.

61. *Hsu D. F. and Shapiro J.* Bounds for the minimal number of transmission delays in double loop networks // *J. Combinat., Inform. Syst. Sci.* 1991. No. 16. P. 55–62.
62. *Hsu D. F. and Shapiro J.* A census of tight one-optimal double loop networks // *Graph Theory, Combinatorics, Algorithms and Applications* / eds. J. Alavi et al. SIAM, 1991. P. 254–265.
63. *Huber K.* Codes over Tori // *IEEE Trans. Inform. Theor.* 1997. V. 43. No. 2. P. 740–744.
64. *Hwang F. K.* A complementary survey on double-loop networks // *Theor. Comp. Sci.* 2001. No. 263. P. 211–229.
65. *Hwang F. K.* A survey on multi-loop networks // *Theor. Comp. Sci.* 2003. No. 299. P. 107–121.
66. *Imase M. and Iton M.* Desing to minimize diameter building-block networks // *IEEE Trans. Comput.* 1981. No. C30. P. 439–442.
67. *Jia X.-D. and Su W.* Triple Loop Networks with Minimal Transmission Delay // *Int. J. Found. Comp. Sci.* 1997. V. 8. No. 3. P. 305–328.
68. *Kotsis G.* Interconnection Topologies and Routing for Parallel Processing Systems // *Austrian—Hungarian Workhop, Technical Report. KFKI*, 1992. P. 95–106.
69. *LaForge L. E., Korver K. F., and Fadali M. S.* What Designers of Bus and Network Architectures Should Know about Hypercubes // *IEEE Trans. Comput.* 2003. V. 52. No. 4. P. 525–544.
70. *Li C. H.* On isomorphisms of finite Cayley graphs — a survey // *Discr. Math.* 2002. V. 256. Issues 1–2. P. 301–334.
71. *Liaw S.-C., Chang G. J., Cao F., and Hsu D. F.* Fault-tolerant Routing in Circulant Networks and Cycle Prefix Networks // *Ann. Comb.* 1998. No. 2. P. 165–172.
72. *Liestman A. L., Opatrny J., and Zaragoza M.* Network Properties of Double and Triple Fixed-Step Graphs // *Int. J. Found. Comp. Sci.* 1998. No. 9. P. 57–76.
73. *Mans B.* On the Interval Routing of Chordal Rings // *Inter. Symposium on Parallel Architectures, Algorithms and Networks (ISPAN 1999)*, 1999, Australia, IEEE Computer Society. P. 16–21.
74. *Mans B., Pappalardi F., and Shparlinski I.* On the Adam Conjecture on Circulant Graphs // *LNCS.* 1998. V. 1449. P. 251–260.
75. *Mans B., Pappalardi F., and Shparlinski I.* On the spectral Adam property for circulant graphs // *Discr. Math.* 2002. V. 254. No. 1–3. P. 309–329.
76. *Mans B. and Shparlinski I.* Bisecting and Gossiping in Circulant Graphs // *LNCS.* 2004. V. 2976. P. 589–598.
77. *Martinez C., Beivide R., Izu C., and Miguel-Alonso J.* Characterization of the Class of Optimal Dence Circulant Graphs of Degree Four // *XIV Jornadas de Paralelismo. Leganes, Septiembre 2003.* P. 1–6.
78. *Martinez C., Beivide R., Stafford E., et al.* Modeling Toroidal Networks with the Gaussian Integers // *IEEE Trans. Comput.* 2008. V. 57. No. 8. P. 1046–1056.
79. *Monakhova E. A.* Optimal circulant computer networks // *Inter. Conf. “Parallel Computing Technologies”, Novosibirsk, USSR; World Scientific, Singapore*, 1991. P. 450–458.
80. *Monakhova E. A.* Algorithms and lower bounds for p-gossiping in circulant networks // *Third Inter. Symposium on Parallel Architectures, Algorithms, and Networks (I-SPAN’97)*, Taipei, Taiwan, Dec. 1997, IEEE Computer Society, Los Alamitos, California. P. 132–137.
81. *Monakhova E. A.* Optimal Triple Loop Networks with Given Transmission Delay: Topological Design and Routing // *Inter. Network Optimization Conference, (INOC’2003)*, Evry/Paris, France. 2003. P. 410–415.

82. *Monakhov O. G. and Monakhova E. A.* Computer Discovery of Analytical Descriptions of Families of Circulant Networks // 6th Inter. Conf. on Soft Computing and Measurements, (SCM'2003), July 25-27, 2003, St.-Petersburg, Russia, 2003. V. 1. P. 345–348.
83. *Monakhov O. G. and Monakhova E. A.* An Algorithm for Discovery of New Families of Optimal Regular Networks // Lect. Notes Artific. Intell. 2003. V. 2843. P. 244–254.
84. *Muga F. P.* Undirected circulant graphs // Inter. Symp. on Parallel Architectures, Algorithms and Networks. IEEE, 1994. P. 113–118.
85. *Muga F. P.* Maximal Order of 3- and 5-Regular Circulant Graphs // Matimyas Matematika. 1999. V. 22. No. 3. 6 p.
86. *Muga F. P. and Yu W. E.* A Proposed Topology for a 192-Processor Symmetric Cluster with a Single-Switch Delay // Proceedings of the First Philippine Computing Science Congress. Manila, Philippines, Nov. 2000. 10 p.
87. *Mukhopadhyaya K. and Sinha B. P.* Fault-tolerant routing in distributed loop networks // IEEE Trans. Comput. 1995. V. 44. No. 12. P. 1452–1456.
88. *Muzychuk M.* On Adam's conjecture for circulant graphs // Discr. Math. 1997. V. 167/168. P. 497–510.
89. *Muzychuk M. E. and Tinhofer G.* Recognizing circulant graphs of prime order in polynomial time // The Electr. J. Combinat. R25. 1998. V. 5. No. 1. P. 501–528.
90. *Narayanan L. and Opatrny J.* Compact routing on chordal rings of degree four // eds. D. Krizanc and P. Widmayer. Sirocco 97: Proc. of the 4th Inter. Colloquium on Structural Information and Communication Complexity, Ascona, Switzerland, Carleton Scientific, 1997. P. 125–137.
91. *Narayanan L., Opatrny J., and Sotteau D.* All-to-All Optical Routing in Chordal Rings of Degree Four // Algorithmica. 2001. V. 31. No. 2. P. 155–178.
92. *Obradovic N., Peters J., and Ruzic G.* Reliable Broadcasting in Double Loop Networks // Networks. 2005. V. 46. No. 2. P. 88–97.
93. *Parhami B.* A Class of Odd-Radix Chordal Ring Networks // J. Comput. Sci. Engin. 2006. V. 4. No. 2–4. P. 1–9.
94. *Parhami B.* Chordal Rings Based on Symmetric Odd-Radix Number Systems // Inter. Conf. on Communications in Computing, Las Vegas, NV, June 27-30, 2006. P. 196–199.
95. *Park J.-H. and Chwa K.-Y.* Recursive Circulant: a New Topology for Multicomputer Networks // Proc. of the Inter. Symp. on Parallel Architectures, Algorithms, and Networks (I-SPAN'94), Kanazawa, Japan, IEEE Computer Society Press, 1994. P. 73–80.
96. *Pelc A.* Fault-Tolerant Broadcasting and Gossiping in Communication Networks // Networks. 1996. No. 28. P. 143–156.
97. *Puente V., Gregorio J.-A., Prellezo J. M., et al.* Adaptive Bubble Router: a Design to Balance Latency and Throughput in Networks for Parallel Computers // Proc. of the 1999 Inter. Conf. on Parallel Processing, ICPP'99, September, 1999. IEEE Computer Society. P. 58–67.
98. *Puente V., Izu C., Gregorio J.-A., et al.* Improving Parallel System Performance by Changing the Arrangement of the Network Links // Intern. Conf. on Supercomputing, May 2000, Santa Fe, New Mexico, USA, ACM, 2000. P. 44–53.
99. *Robič B.* Optimal routing in 2-jump circulant networks // Tech. Report N397. Cambridge: University of Cambridge Computer Laboratory, 1996. 7 p.
100. *Robič B. and Žerovnik J.* Minimum 2-terminal routing in 2-jump circulant graphs // Comput. Artific. Intell. 2000. V. 19. No. 1. P. 37.
101. *Sampels M.* Cayley graphs as interconnection networks: A case study // Inter. Conf. Parcella'96. Berlin: Akademie Verlag, 1996. P. 67–76.

102. *Sampels M.* Large networks with small diameter // Inter. Workshop on Graph Theoretic Concepts in Computer Science (WG'97). Berlin: Springer, 1997. P. 288–302.
103. *Schinder M.* New architectures keep pace with throughput needs // Electr. Design. 1981. No. 5. P. 97–106.
104. *Shin K. G.* HARTS: A Distributed Real-Time Architecture // Computer. 1991. V. 24. No. 5. P. 25–35.
105. *Stojmenovic I.* Multiplicative circulant networks. Topological properties and communication algorithms // Discr. Appl. Math. 1997. No. 77. P. 281–305.
106. *Stone H. S.* The organization of high-speed memory for parallel block transfer data // IEEE Trans. Comput. 1970. No. 19. P. 47–53.
107. *Toueg S. and Steiglitz K.* The desing of small diameter networks by local search // IEEE Trans. Comput. 1979. No. 28. P. 537–542.
108. *Turner J.* Point-symmetric graphs with a prime number of points // J. Combin. Theory. 1967. No. 3. P. 136–145.
109. *Tzvieli D.* Minimal diameter double-loop networks. 1. Large infinite optimal families // Networks. 1991. No. 21. P. 387–415.
110. *Wilkov R. S.* Analysis and design of reliable computer networks // IEEE Trans. Comput. 1972. V. 20. No. 3. P. 660–678.
111. *Wong C. K. and Coppersmith D.* A combinatorial problem related to multimodule memory organizations // J. Assoc. Comput. Mach. 1974. No. 21. P. 392–402.
112. *Wong C. K. and Maddocks T. W.* A generalized Pascal's triangle // Fibonacci Quart. 1975. No. 13. P. 134–136.
113. *Yang Y., Funashashi A., Jouraku A., et al.* Recursive Diagonal Torus: An Interconnection Network for Massively Parallel Computers // IEEE Trans. Parallel and Distributed Systems. 2001. V. 12. No. 7. P. 701–715.
114. *Yebra J. L. A., Fiol M. A., Morillo P., and Alegre I.* The diameter of undirected graphs associated to plane tessellations // Ars Combinat. 1985. No. 20B. P. 159–172.
115. *Žerovnik J. and Pisanski T.* Computing the diameter in multiple-loop networks // J. Algorithms. 1993. No. 14. P. 226–243.

ВЫЧИСЛИТЕЛЬНЫЕ МЕТОДЫ В ДИСКРЕТНОЙ МАТЕМАТИКЕ

DOI 10.17223/20710410/13/9

УДК 519.6

О РЕАЛИЗАЦИИ МЕТОДА СОГЛАСОВАНИЯ В КРИПТОАНАЛИЗЕ С ПОМОЩЬЮ ПАРАЛЛЕЛЬНЫХ ВЫЧИСЛЕНИЙ¹

В. М. Фомичев

*Институт проблем информатики РАН, г. Москва, Россия***E-mail:** fomichev@nm.ru

Оценено время реализации метода согласования применительно к анализу итеративных симметричных блочных шифров с использованием кластерных и распределенных вычислений. Показано, что по сравнению с однопроцессорной системой коэффициент сокращения времени может достигать числа используемых процессоров.

Ключевые слова: метод согласования, кластер, распределенные вычисления.

Введение

Метод согласования [1], или атака «встреча посередине» (meet-in-the-middle attack) [2–5], применяется для определения ключа шифра, как правило, по известным открытому и зашифрованному текстам. Он менее трудоемок по сравнению с полным опробованием ключей, если функция шифрования $E(q, x)$ открытого текста x по ключу $q \in V_n = \{0, 1\}^n$ допускает декомпозицию на две функции как $E(q, x) = g(q, g'(q, x))$, где для множеств существенных ключевых переменных K и K' соответственно функций g и g' выполнено $K \setminus K' \neq \emptyset$ и $K' \setminus K \neq \emptyset$. Наибольший эффект от применения метода достигается, если множества K и K' равномощны и $K \cap K' = \emptyset$. При этом опробование ключа выполняется как независимое опробование переменных из множеств K и K' и ключ q определяется с вычислительной сложностью порядка $O(2^{n/2})$ операций типа зашифрования-расшифрования при использовании памяти, достаточной для хранения порядка $O(2^{n/2})$ ключей.

Параллельные вычисления с использованием N процессоров позволяют сократить в N раз время решения некоторых задач, например полного опробования ключей. Вместе с тем применение параллельных вычислений для решения других задач не столь эффективно. Оценим эффективность различных моделей параллельных вычислений для реализации метода согласования в решении следующей задачи.

Для r -раундового симметричного блочного шифра требуется вычислить n -битовый ключ q по известным t -битовым блокам x и y открытого и зашифрованного текстов, где $t \geq n$ и ключ q по блокам x и y определяется однозначно, при следующих предположениях:

¹Работа выполнена в рамках мероприятия 1.2.1 ФЦП «Научные и научно-педагогические кадры инновационной России» на 2009–2013 гг. по направлению «Распределенные вычислительные системы».

- 1) ключ q есть конкатенация независимых ключей: $q = v \cdot w$, где $v \in V_m$, $w \in V_{n-m}$ и $m \leq n/2$;
- 2) функции шифрования первых $l < r$ раундов и остальных $r - l$ раундов шифрования суть подстановки соответственно g_v и z_w , определяемые бинарными ключами v и w .

Тогда зашифрование блока x в блок y с помощью подстановки E_q имеет вид

$$y = E_q(x) = g_v z_w(x) = z_w(g_v(x)). \quad (1)$$

Корректность предложенных ниже алгоритмов следует из (1).

Модель вычислительной системы предполагает использование N идентичных вычислителей с неограниченной памятью (размер памяти далее уточняется), где у вычислителей одинаковы производительность вычислений, скорости чтения/записи данных и др. Различаются два случая: кластерные вычисления (КВ) и распределенные вычисления (РВ). Преимущество модели КВ заключается в возможности достаточно активного обмена данными между вычислителями. Преимущество модели РВ, где координатор распределяет задания между процессорами — участниками вычислений и объединяет результаты вычислений, в том, что число участников РВ может заметно превышать число процессоров в кластере. При этом участник обменивается данными только с координатором.

1. Кластерные вычисления

Пусть кластерная система имеет 2^k вычислителей, снабженных блоками памяти, $k \leq m$. Каждый вычислитель имеет номер, являющийся его адресом (число от 0 до $2^k - 1$, или в двоичной записи — вектор из V_k). Между блоками памяти вычислителей может выполняться интенсивный обмен данными. Для реализации алгоритма каждый вычислитель использует адресную память размера 2^{t-k} ячеек, в ячейке могут быть записаны несколько вариантов ключей, то есть элементов V_n . Адреса ячеек суть элементы V_{t-k} , являющиеся значениями некоторой хеш-функции: $V_t \rightarrow V_{t-k}$. Хеш-функция может быть весьма простой, например выделение первых $t - k$ битов двоичного t -битового вектора.

Для любого двоичного вектора $(\alpha_1, \alpha_2, \dots)$ размерности больше k обозначим

$$\delta(\alpha_1, \alpha_2, \dots) = (\alpha_1, \dots, \alpha_k), \quad \bar{\delta}(\alpha_1, \alpha_2, \dots) = (\alpha_{k+1}, \alpha_{k+2}, \dots).$$

Для вектора $\alpha = (\alpha_1, \dots, \alpha_k) \in V_k$ и пространства векторов V_s , где $s \geq k$, обозначим

$$V_s(\alpha) = \{\xi \in V_s : \delta(\xi) = \alpha\}.$$

Алгоритм состоит из предварительного и оперативного этапов.

Предварительный этап (заполнение блоков памяти вычислителей)

Вычислитель с номером $\alpha \in V_k$ последовательно опробует ключи v из $V_m(\alpha)$ и вычисляет $g_v(x)$ для блока x . Затем пара $(v, g_v(x))$ направляется вычислителю с номером $\delta(g_v(x))$, который записывает ключ v в свою память по адресу $\bar{\delta}(g_v(x))$.

По завершении этапа множество ключей из V_m распределено по ячейкам памяти всех вычислителей. Обозначим через $Q(\alpha, \beta)$ множество ключей из V_m , записанных в памяти вычислителя с номером α по адресу β .

Оперативный этап (определение ключа)

- 1) Вычислитель с номером $\alpha \in V_k$ последовательно опробует ключи w из $V_{n-m}(\alpha)$ и вычисляет $(z_w)^{-1}(y)$ для блока y , затем пара $(w, (z_w)^{-1}(y))$ направляется вычислителю с номером $\delta((z_w)^{-1}(y))$.

- 2) Вычислитель с номером $\delta((z_w^{-1})(y))$ обращается в свою память по адресу $\bar{\delta}((z_w^{-1})(y))$. Конкатенация $v \cdot w$ для каждого ключа v из множества $Q = Q(\delta((z_w^{-1})(y)), \bar{\delta}((z_w^{-1})(y)))$ есть кандидат на значение искомого ключа $q = v \cdot w$. Если $Q \neq \emptyset$, то выполняется отбраковка всех ключей вида $v \cdot w$ (например, по критерию соответствия известным парам открытого и зашифрованного текстов).

Характеристики метода

Оценим (в предположении, что ключ q выбирается случайно равномерно из множества V_n) среднее время $T(m)$ описанной реализации метода согласования через время реализации операций зашифрования, расшифрования, пересылки и обращения в память, обозначаемых соответственно τ_3 , τ_p , τ_n , τ_o . Положим, что работа алгоритма происходит в дискретные моменты времени (такты) и в каждый такт на первом этапе в любую ячейку памяти записывается не более одного варианта ключа v , на втором этапе из любой ячейки памяти извлекается не более одного варианта ключа v , т. е. замедления «из-за очередей» в работе вычислителей не происходит.

На первом этапе для ключа v из $V_m(\alpha)$ реализуется однократное зашифрование, пересылка и запись в память, отсюда среднее время $T_1(m)$ работы вычислителя равно $2^{m-k}(\tau_3 + \tau_n + \tau_o)$. На втором этапе для ключа w из $V_{n-m}(\alpha)$ реализуется однократное расшифрование, пересылка и обращение в память. Следовательно, среднее время $T_2(m)$ выполнения вычислителем второго этапа равно $2^{n-m-k}(\tau_p + \tau_n + \tau_o) + T_{бр}$, где $T_{бр}$ — среднее время отбраковки кандидатов на значение ключа.

Оценим величину $T_{бр}$. В каждой ячейке памяти записано в среднем 2^{m-t} вариантов ключа v . Среднее число обращений в любую ячейку памяти на втором этапе равно 2^{n-m-t} . Отсюда каждым вычислителем отбраковывается в среднем 2^{n-t-k} кандидатов на значение ключа. Значит, $T_{бр} = 2^{n-t-k}\tau_3$, и отбраковка кандидатов на значение ключа вносит несущественный вклад в общую трудоемкость. Следовательно,

$$T(m) = T_1(m) + T_2(m) \approx 2^{m-k}(\tau_3 + \tau_n + \tau_o) + 2^{n-m-k}(\tau_p + \tau_n + \tau_o).$$

Отсюда, если $\tau_3 \approx \tau_p$, то минимум трудоёмкости $T(m)$ достигается при $m = \lfloor n/2 \rfloor$:

$$T = T(\lfloor n/2 \rfloor) \approx 2^{n/2-k+1}(\tau_3 + \tau_n + \tau_o). \quad (2)$$

Следовательно, среднее время T оценивается величиной порядка $O(\tau 2^{n/2-k})$, где $\tau = \max\{\tau_3, \tau_n, \tau_o\}$.

Надёжность метода равна 1. В связи с минимизацией по m трудоемкости $T(m)$ уточним размер требуемой памяти: вычислителю достаточно иметь $2^{n/2-k}$ ячеек, в которые записываются элементы $V_{n/2}$. Адресами ячеек являются элементы $V_{n/2-k}$.

При выборе оптимального (по времени реализации алгоритма) размера памяти следует учесть, что реализация алгоритма может замедляться «из-за очередей», когда в одну ячейку одновременно поступает несколько запросов в связи с необходимостью записи или извлечения информации. Это замедление тем несущественней, чем меньше соотношение τ_o/τ_3 .

Таким образом, время определения ключа блочного шифра методом согласования с использованием кластерных вычислений с числом процессоров 2^k может быть сокращено до 2^k раз по сравнению с однопроцессорной вычислительной системой, если время пересылки данных между вычислителями не слишком велико. Важно также, что совокупный объем требуемой памяти $2^{n/2}$ ячеек также распределяется между 2^k процессорами.

2. Распределенные вычисления

В системе РВ с 2^p участниками (вычислителями), $p \leq m$, каждый участник имеет номер, являющийся его адресом (число от 0 до 2^{p-1} , или в двоичной записи — вектор из V_p). Алгоритм использует 2^t ячеек адресной памяти координатора (адрес ячейки есть элемент V_t), в каждую из них могут быть записаны несколько вариантов ключей — элементов V_n . Участники могут отправлять данные координатору, но не могут обращаться к его памяти.

Алгоритм состоит из предварительного и оперативного этапов.

Предварительный этап (заполнение памяти координатора)

Вычислитель с номером $\alpha \in V_p$ последовательно при каждом ключе v из $V_m(\alpha)$ вычисляет $g_v(x)$ для блока x и направляет пару $(v, g_v(x))$ координатору, где ключ v записывается в память координатора по адресу $g_v(x)$. По завершении этапа множество ключей из V_m распределено по ячейкам памяти координатора. Обозначим через $Q(\beta)$ множество ключей из V_m , записанных в памяти координатора по адресу β .

Оперативный этап (определение ключа)

- 1) Вычислитель с номером $\alpha \in V_p$ последовательно при каждом ключе w из $V_{n-m}(\alpha)$ вычисляет $(z_w^{-1})(y)$ для блока y и направляет пару $(w, (z_w^{-1})(y))$ координатору.
- 2) Координатор обращается в память по адресу $(z_w^{-1})(y)$. Конкатенация каждого ключа v из $Q((z_w^{-1})(y))$ с ключом w есть кандидат на значение искомого ключа $q = v \cdot w$. Если $Q((z_w^{-1})(y)) \neq \emptyset$, то координатор подвергает отбраковке все ключи вида $v \cdot w$ (например, по критерию соответствия известным парам открытого и зашифрованного текстов).

Характеристики метода

Оценим (ключ q выбирается случайно равновероятно из V_n) среднее время $T(m)$ описанной реализации метода согласования через время реализации операций зашифрования, расшифрования, пересылки и обращения в память. Положим, что в каждый такт на первом этапе в любую ячейку памяти записывается не более одного варианта ключа v , на втором этапе из любой ячейки памяти извлекается не более одного варианта ключа v , т. е. замедления «из-за очередей» в работе вычислителей не происходит.

Среднее время $T_{1y}(m)$ выполнения первого этапа участниками равно $2^{m-p}(\tau_3 + \tau_{\Pi})$, так как для каждого ключа v из $V_m(\alpha)$ реализуется по одной операции зашифрования и пересылки. Среднее время $T_{1к}(m)$ выполнения первого этапа координатором (записи в память) равно $2^m \tau_0$.

Следовательно, среднее время $T_1(m)$ выполнения первого этапа равно

$$T_1(m) = T_{1y}(m) + T_{1к}(m) = 2^{m-p}(\tau_3 + \tau_{\Pi} + 2^p \tau_0).$$

Для ключа w из $V_{n-m}(\alpha)$ участником реализуется по одной операции расшифрования и пересылки. Следовательно, среднее время $T_{2y}(m)$ выполнения второго этапа каждым участником равно $2^{n-m-p}(\tau_p + \tau_{\Pi})$. Среднее время $T_{2к}(m)$ выполнения второго этапа координатором (записи в память и отбраковки) равно $2^{n-m} \tau_0 + T_{бр}$, где $T_{бр}$ — среднее время отбраковки кандидатов на значение ключа. Отсюда среднее время $T_2(m)$ выполнения второго этапа равно $2^{n-m-p}(\tau_p + \tau_{\Pi} + 2^p \tau_0) + T_{бр}$.

Оценим величину $T_{бр}$. В каждой ячейке памяти записано в среднем 2^{m-t} вариантов ключа v . Среднее число обращений в память на втором этапе равно 2^{n-m} . Отсюда координатором отбраковывается в среднем 2^{n-t} кандидатов на значение ключа. Значит, $T_{бр} = 2^{n-t} \tau_3$, и отбраковка кандидатов в ключи вносит несущественный вклад в общую трудоемкость. Следовательно,

$$T(m) = T_1(m) + T_2(m) \approx 2^{m-p}(\tau_3 + \tau_{\Pi} + 2^p\tau_0) + 2^{n-m-p}(\tau_p + \tau_{\Pi} + 2^p\tau_0).$$

Отсюда, если $\tau_3 \approx \tau_p$, то минимум среднего времени $T(m)$ достигается при $m = \lfloor n/2 \rfloor$:

$$T = T(\lfloor n/2 \rfloor) \approx 2^{n/2-p}(\tau_3 + \tau_{\Pi} + 2^p\tau_0). \quad (3)$$

В связи с минимизацией $T(m)$ по m уточним размер требуемой памяти: координатору достаточно иметь $2^{n/2}$ ячеек, в которые записываются элементы $V_{n/2}$. Адресами ячеек являются элементы $V_{n/2}$. Следовательно, среднее время работы алгоритма согласования по сравнению с полным опробованием ключей сокращается не более чем в 2^p раз, и сокращение зависит от соотношения величин τ_0 и $\max\{\tau_3, \tau_{\Pi}\}$. Надёжность метода равна 1.

3. Комбинирование кластерных и распределенных вычислений

В данной модели РВ система использует 2^p участников, $p \leq m$. Каждый участник имеет номер, являющийся его адресом (число от 0 до 2^{p-1} , или в двоичной записи — вектор из V_p). Координатор располагает кластерной подсистемой 2^k вычислителей, $k \leq p$, каждый вычислитель кластерной системы имеет номер, являющийся его адресом (число от 0 до $2^k - 1$, или в двоичной записи — вектор из V_k), и имеет блок памяти размера 2^{t-k} ячеек (адрес ячейки есть элемент V_{t-k}). В каждую ячейку могут быть записаны несколько вариантов ключей — элементов V_n . Участники РВ могут отправлять данные кластерным вычислителям, но не могут обращаться в память кластерных вычислителей.

Предварительный этап (заполнение памяти координатора)

Участник с номером $\alpha \in V_p$ последовательно при каждом ключе v из $V_m(\alpha)$ вычисляет $g_v(x)$ для блока x и направляет пару $(v, g_v(x))$ кластерному вычислителю с номером $\delta(g_v(x))$, который записывает в память ключ v по адресу $\bar{\delta}(g_v(x))$. По завершении этапа множество ключей из V_m распределено по блокам памяти кластерных вычислителей координатора. Обозначим через $Q(\alpha, \beta)$ множество ключей из V_m , записанных в блоке памяти вычислителя с номером α по адресу β .

Оперативный этап (определение ключа)

- 1) Вычислитель с номером $\alpha \in V_p$ последовательно при каждом ключе w из $V_{n-m}(\alpha)$ вычисляет $(z_w^{-1})(y)$ для блока y и направляет пару $(w, (z_w^{-1})(y))$ кластерному вычислителю с номером $\delta((z_w^{-1})(y))$.
- 2) Кластерный вычислитель с номером $\delta((z_w^{-1})(y))$ вычисляет адрес $\bar{\delta}((z_w^{-1})(y))$ и обращается к своему блоку памяти по этому адресу. Конкатенация вида $v \cdot w$ для каждого ключа v из множества $Q = Q(\delta((z_w^{-1})(y)), \bar{\delta}((z_w^{-1})(y)))$ есть кандидат на значение ключа. Если $Q \neq \emptyset$, то вычислитель с номером $\delta((z_w^{-1})(y))$ все ключи вида $v \cdot w$ подвергает отбраковке (например, по критерию соответствия известным парам открытого и зашифрованного текстов).

Характеристики метода

Оценим (ключ q выбирается случайно равномерно из V_n) среднее время $T(m)$ описанной реализации метода согласования через время реализации операций зашифрования, расшифрования, пересылки и обращения в память.

Среднее время $T_{1y}(m)$ выполнения первого этапа участником равно $2^{m-p}(\tau_3 + \tau_{\Pi})$, так как для каждого ключа v из $V_m(\alpha)$ реализуется по одной операции зашифрования и пересылки. Среднее время $T_{1k}(m)$ выполнения первого этапа кластерным вычислителем оценивается величиной $2^{m-k}\tau_0$, так как запись в память 2^m ключей v выполняется

2^k вычислителями. Отсюда среднее время выполнения первого этапа алгоритма определяется величиной $\max\{2^{m-p}(\tau_3 + \tau_{\Pi}), 2^{m-k}\tau_o\}$.

Для ключа w из $V_{n-m}(\alpha)$ участником реализуется одно расшифрование и пересылка. Следовательно, среднее время $T_{2_y}(m)$ выполнения второго этапа участником равно $2^{n-m-p}(\tau_p + \tau_{\Pi})$. Среднее время $T_{2_k}(m)$ выполнения второго этапа кластерным вычислителем оценивается величиной $2^{n-m-k}\tau_o + T_{6p}$, так как запись в память 2^{n-m} ключей w выполняется 2^k вычислителями. Среднее время T_{6p} отбраковки кандидатов на значение искомого ключа вносит несущественный вклад в общую трудоемкость. Отсюда среднее время выполнения второго этапа определяется величиной $\max\{2^{n-m-p}(\tau_p + \tau_{\Pi}), 2^{n-m-k}\tau_o\}$. Следовательно, время $T(m)$ определяется величиной порядка

$$\max\{2^{m-p}(\tau_3 + \tau_{\Pi}), 2^{n-m-p}(\tau_p + \tau_{\Pi}), 2^{m-k}\tau_o, 2^{n-m-k}\tau_o\}. \quad (4)$$

Тогда минимум $T(m)$ достигается при $m = \lfloor n/2 \rfloor$ и при $\tau_3 = \tau_p$ верны оценки

$$O((\tau_3 + \tau_{\Pi})2^{n/2-p}) \leq \min T(m) \leq O(\tau_o 2^{n/2-k}).$$

Следовательно, $\min T(m)$ может быть сокращен в несколько раз по сравнению с КВ и РВ (ср. с формулами (2), (3)). Коэффициент сокращения определяется соотношением скоростей шифрования, пересылки данных и обращения к памяти. Надёжность метода равна 1.

Уточним размер требуемой памяти: кластерному вычислителю достаточно иметь $2^{n/2-k}$ ячеек, в которые записываются элементы $V_{n/2}$. Адресами ячеек являются элементы $V_{n/2-k}$.

Выводы

Время определения ключа блочного шифра методом согласования может быть существенно сокращено по сравнению с однопроцессорной вычислительной системой:

- 1) при использовании КВ с числом процессоров 2^k — примерно в 2^k раз;
- 2) при использовании РВ с 2^p участниками — до 2^p раз, сокращение определяется соотношением скоростей шифрования, пересылки данных и обращения к памяти координатора;
- 3) при использовании РВ с 2^p участниками и подсистемы КВ координатора с числом процессоров 2^k , где $k \leq p$ — от 2^k до 2^p раз, сокращение определяется соотношением скоростей шифрования, пересылки данных и обращения к памяти кластерных вычислителей.

При использовании КВ память распределяется по вычислителям кластерной системы.

ЛИТЕРАТУРА

1. Фомичёв В. М. Методы дискретной математики в криптологии. М.: ДИАЛОГ-МИФИ, 2010.
2. Шнайер Б. Прикладная криптография. Протоколы, алгоритмы, исходные тексты на языке Си. М.: ТРИУМФ, 2002.
3. Словарь криптографических терминов / под ред. Б. А. Погорелова и В. Н. Сачкова. М.: МЦНМО, 2006. 94 с.
4. Брассар Ж. Современная криптология: пер. с англ. М.: Полимед, 1999. 173 с.
5. Грушо А. А., Тимонина Е. Е., Применко Э. А. Анализ и синтез криптоалгоритмов. Курс лекций. Йошкар-Ола: МФ МОСУ, 2000.

МАТЕМАТИЧЕСКИЕ ОСНОВЫ ИНТЕЛЛЕКТУАЛЬНЫХ СИСТЕМ

DOI 10.17223/20710410/13/10

УДК 519.168

АНАЛИТИЧЕСКИЙ МЕТОД ДООПРЕДЕЛЕНИЯ КРАТНЫХ ПРЕДПОЧТЕНИЙ В МАТРИЦЕ ПАРНЫХ СРАВНЕНИЙ

И. С. Киселев

*Петербургский государственный университет путей сообщения, г. Санкт-Петербург,
Россия*

E-mail: igor@kiselev.spb.ru

Предложен аналитический метод доопределения кратных предпочтений в матрице парных сравнений, имеющий полиномиальную сложность от размерности матрицы и количества доопределяемых элементов, не ухудшающий показатель согласованности исходной матрицы.

Ключевые слова: матрица парных сравнений, согласованность.

Введение

Матрица парных сравнений получила широкое применение для вычисления приоритетов сущностей на основе предпочтений экспертов. Одним из распространенных типов предпочтений является кратность превосходства. Рассмотрим матрицу кратности парных сравнений A . В ней элемент a_{ij} определяет, во сколько раз сущность x_i предпочтительнее сущности x_j . Данный тип предпочтения лег в основу метода анализа иерархий (МАИ) Т. Саати [1]. Для получения вектора весов оцениваемых объектов в МАИ находится нормированный собственный вектор матрицы парных сравнений, соответствующий ее максимальному собственному числу. В идеально согласованной матрице экспертных предпочтений для каждой тройки элементов выполняется равенство $a_{ij} \cdot a_{jk} = a_{ik}$. Функция оценивания согласованности предпочтений убывает с увеличением числа троек, для которых это условие не выполняется, и с ростом разности $|a_{ij} \cdot a_{jk} - a_{ik}|$.

Трудоёмкость формирования матрицы кратности предпочтений быстро растёт с ростом её размерности, отсюда возникает задача доопределения элементов матриц при частичном задании предпочтений [2–5]. В. Д. Ногин [2, 3] рассмотрел доопределение идеально согласованных матриц; В. Г. Тоценко [4] — доопределение матриц с внесением изменений в исходно заданные оценки. В работе [5] построение проводится с использованием целевой функции, основанной на индексе согласованности (ИС), предложенном Т. Саати. Однако ИС имеет искусственный характер, и в настоящей работе предлагается аналитическое решение задачи с использованием функции согласованности, имеющей физический смысл.

1. Постановка задачи

В [5] для определения оптимальной величины кратности a_{ij}^{opt} предпочтения $x_i \succ x_j$ используется целевая функция

$$\eta_\lambda(a_{ij}) = (N - 1 + a_{ii}) / \lambda_{\max} \rightarrow \max. \quad (1)$$

В качестве ограничений задачи оптимизации принимаются все известные значения предпочтений из матрицы A . Формула (1) связывает максимальное собственное число λ_{\max} с размерностью матрицы $N \times N$ согласно введённому Саати индексу согласованности ИС [1]

$$\text{ИС} = \frac{\lambda_{\max} - N}{N - 1}.$$

Незаполненные клетки в предъявленной для доопределения матрице кратности предпочтений заполняются единицами. Установлено, что функция $\eta_{\lambda}(a_{ij})$ дифференцируема и имеет единственный максимум в диапазоне $(0, \infty)$. Оптимальное значение кратности a_{ij}^{opt} находится путем итеративного пересчёта по формуле

$$a_{ij}^* = \arg(\max(\eta_{\lambda}(a_{ij}))). \quad (2)$$

Нахождение максимального собственного числа матрицы является трудоёмким процессом, поэтому была предложена целевая функция $\eta_D(a_{ij}) \rightarrow \max$, основанная на нахождении определителя D матрицы A . Однако в работе [6] на основе имитационного моделирования показано, что определитель матрицы даёт решение задачи оптимизации не во всех случаях.

Для сравнения оценок согласованности матриц различных размерностей вводится отношение согласованности (ОС), рассчитываемое как отношение ИС к случайной согласованности (СС) ($\text{ОС} = \text{ИС}/\text{СС}$). СС — это математическое ожидание ИС для матрицы требуемой размерности, заполненной случайными величинами. Для практических расчетов СС берется из таблицы, приведенной в [1]. Показано [7], что определение ИС через соотношение максимального положительного собственного числа матрицы с её размерностью имеет искусственный характер. Это подтверждается ограниченной областью использования вычисляемого на его основе ОС, поскольку для плохо согласованной матрицы значение ОС выходит за рамки шкалы $[0, 1]$.

Таким образом, с одной стороны, определение оптимальных значений неизвестных предпочтений по формуле (2) является трудоёмким процессом, а с другой — сама целевая функция (1) имеет искусственный характер.

Граф, матрица связности которого совпадает с рассматриваемой матрицей парных сравнений, будем называть графом, соответствующим матрице. Для оценки качественной несогласованности достаточно подсчитать количество циклов в этом графе. В случае количественных предпочтений для получения оценки их несогласованности этого недостаточно, так как каждая тройка сущностей содержит не только факт согласованности, но и ее меру. Для получения соответствующего показателя s матрицы парных сравнений будем суммировать меру согласованности всех троек сущностей [7]:

$$s = \sum_{i=1}^{N-2} \sum_{j=i+1}^{N-1} \sum_{k=j+1}^N f(a_{ij}, a_{jk}, a_{ki}). \quad (3)$$

За меру согласованности предпочтений тройки сущностей примем отклонение от условия идеальной согласованности $a_{ij} \cdot a_{jk} \cdot a_{ik}^{-1} = a_{ij} \cdot a_{jk} \cdot a_{ki} = 1$. Для перехода от мультипликативной к аддитивной форме представления условия согласованности воспользуемся логарифмической шкалой, а для получения неотрицательного значения меры согласованности возьмём квадрат от логарифма произведения. В результате функция (3) примет следующий вид:

$$s = \sum_{i=1}^{N-2} \sum_{j=i+1}^{N-1} \sum_{k=j+1}^N \ln^2(a_{ij} \cdot a_{jk} \cdot a_{ki}). \quad (4)$$

Данный показатель позволяет заменить трудоемкий процесс вычисления максимального собственного числа более простым расчётом.

Решим задачу доопределения одного или нескольких элементов матрицы кратности предпочтений аналитически, используя функцию (4) для вычисления оптимальных значений неизвестных предпочтений, и оценим вычислительную сложность полученного решения.

2. Доопределение одного элемента матрицы

Примем за целевую функцию минимизацию коэффициента несогласованности (4). Коэффициент несогласованности s , являясь функцией от матрицы, является также функцией от всех элементов матрицы. При решении задачи доопределения матрицы все известные элементы являются константами. Тогда можно рассматривать s как функцию только от неизвестных элементов матрицы, а в случае единственного неизвестного элемента — как функцию от одного аргумента. Оптимальное значение неизвестного элемента a_{vw} определяется как $a_{vw}^* = \arg(\min(s(a_{vw}))$). Экстремальное значение функции несогласованности предпочтений имеет место при равенстве её производной нулю: $s' = 0$.

Функция $s(a_{vw})$ на области определения $(0, \infty)$ дифференцируема и имеет единственный экстремум в своей минимальной точке. Значит, для поиска оптимального значения элемента необходимо решить уравнение $\frac{\partial s(a_{vw})}{\partial a_{vw}} = 0$. Возьмем производную от s (4):

$$\begin{aligned} \frac{\partial \sum_{i=1}^{N-2} \sum_{j=i+1}^{N-1} \sum_{k=j+1}^N \ln^2(a_{ij} \cdot a_{jk} \cdot a_{ki})}{\partial a_{vw}} &= \frac{\partial \sum_{\substack{k=1 \\ k \neq v,w}}^N \ln^2(a_{vw} \cdot a_{wk} \cdot a_{kv})}{\partial a_{vw}} = \\ &= \frac{2 \sum_{\substack{k=1 \\ k \neq v,w}}^N \ln(a_{vw} \cdot a_{wk} \cdot a_{kv})}{a_{vw}} = \frac{2 \ln \prod_{\substack{k=1 \\ k \neq v,w}}^N (a_{vw} \cdot a_{wk} \cdot a_{kv})}{a_{vw}} = 0. \end{aligned}$$

Найдем положительное решение данного уравнения относительно a_{vw} :

$$a_{vw} = \sqrt[N-2]{\frac{1}{\prod_{\substack{k=1 \\ k \neq v,w}}^N (a_{wk} \cdot a_{kv})}}.$$

Так как для матрицы парных сравнений кратности предпочтений $a_{vw} = a_{wv}^{-1}$, то

$$a_{wv} = \sqrt[N-2]{\prod_{\substack{k=1 \\ k \neq v,w}}^N (a_{wk} \cdot a_{kv})}.$$

Таким образом, оптимальное значение предпочтения a_{vw} в матрице является средним геометрическим всех путей длины 2 из вершины w в вершину v на графе, соответствующем оптимизируемой матрице.

3. Доопределение нескольких элементов матрицы

Введем сквозную нумерацию элементов верхнего треугольника матрицы слева направо по столбцам с помощью функции нумерации

$$r(i, j) = (j(j - 3))/2 + i + 1, \quad i < j.$$

Например, для матрицы порядка 5 нумерация будет следующей:

	1	2	3	4	5
1	—	1	2	4	7
2	—	—	3	5	8
3	—	—	—	6	9
4	—	—	—	—	10
5	—	—	—	—	—

Будем решать задачу доопределения M элементов матрицы. Пусть неизвестными являются элементы $a_{i_1 j_1}, a_{i_2 j_2}, \dots, a_{i_M j_M}$, где $i_l < j_l$ для всех $l = 1, 2, \dots, M$.

Для нахождения оптимальных значений нескольких элементов матрицы следует найти экстремум функции согласованности матрицы от всех оптимизируемых элементов. Необходимым условием экстремума является равенство нулю всех частных производных функции. Задача оптимизации M элементов сводится к решению системы из M уравнений вида

$$\begin{cases} \frac{\partial s(a_{i_1 j_1}, a_{i_2 j_2}, \dots, a_{i_M j_M})}{\partial a_{i_1 j_1}} = 0, \\ \frac{\partial s(a_{i_1 j_1}, a_{i_2 j_2}, \dots, a_{i_M j_M})}{\partial a_{i_2 j_2}} = 0, \\ \dots \\ \frac{\partial s(a_{i_1 j_1}, a_{i_2 j_2}, \dots, a_{i_M j_M})}{\partial a_{i_M j_M}} = 0, \end{cases}$$

где s — показатель несогласованности (4).

Аналогично случаю оптимизации одного элемента каждое из M уравнений системы примет вид

$$\frac{\partial \sum_{\substack{k=1 \\ k \neq i_l \\ k \neq j_l}}^N \ln^2(a_{i_l j_l} \cdot a_{j_l k} \cdot a_{k i_l})}{\partial a_{i_l j_l}} = 0, \quad l = 1, \dots, M. \tag{5}$$

Преобразуем систему (5):

$$a_{i_l j_l}^{N-2} \prod_{\substack{k=1 \\ k \neq i_l \\ k \neq j_l}}^N (a_{j_l k} \cdot a_{k i_l}) = 1, \quad l = 1, \dots, M. \tag{6}$$

Левая часть каждого из уравнений системы является произведением степеней элементов матрицы. Используя равенство $a_{ij} = a_{ji}^{-1}$, перейдем к элементам только верхней треугольной матрицы. Степени элементов для l -го уравнения будут равны:

- $N - 2$ для $a_{i_l j_l}$;
- 1 для $a_{j_l k}$, $k > j_l$, и для $a_{k i_l}$, $k < i_l$;
- -1 для $a_{k j_l}$, $k < j_l$, $k \neq i_l$, и для $a_{i_l k}$, $k > i_l$, $k \neq j_l$;
- 0 в остальных случаях.

Пусть матрица D содержит степени элементов a_{ij} в уравнениях (6), а именно: элемент матрицы d_{st} , где $s = r(i_l, j_l)$, $t = r(i, j)$, равен степени, в которой a_{ij} входит в l -е уравнение системы.

Перенесем все известные элементы a_{ij} в правые части уравнений (6) и заменим элементы нижней треугольной матрицы их парами из верхней:

$$\prod_{t=1}^M a_{i_t j_t}^{d_{r(i_l, j_l) r(i_t, j_t)}} = \prod_{\substack{(i, j) \neq (i_t, j_t), \\ i < j, t=1, \dots, M}} a_{ij}^{-d_{r(i_l, j_l) r(i, j)}}, \quad l = 1, \dots, M.$$

Для решения этой системы составим матрицу P , состоящую из степеней неизвестных: $p_{st} = d_{r(i_s, j_s) r(i_t, j_t)}$, $s = 1, \dots, M$, $t = 1, \dots, M$.

Тогда решением системы является

$$a_{i_l j_l} = \prod_{k=1}^M \left(\prod_{\substack{(i, j) \neq (i_t, j_t), \\ i < j, t=1, \dots, M}} a_{ij}^{-d_{r(i_k, j_k) r(i, j)}} \right)^{(P^{-1})_{lk}}, \quad l = 1, \dots, M. \quad (7)$$

Пример. Метод будем иллюстрировать на матрице порядка 5 с тремя неизвестными элементами ($N = 5$, $M = 3$):

$$A = \begin{vmatrix} 1 & 4 & a_{13} & 2 & 6 \\ 1/4 & 1 & a_{23} & 1/2 & 4/3 \\ a_{13}^{-1} & a_{23}^{-1} & 1 & 1/5 & a_{35} \\ 1/2 & 2 & 5 & 1 & 3 \\ 1/6 & 3/4 & a_{35}^{-1} & 1/3 & 1 \end{vmatrix}.$$

Матрица D имеет следующий вид:

$$D = \begin{array}{c|cccccccccc} & 12 & 13 & 23 & 14 & 24 & 34 & 15 & 25 & 35 & 45 \\ \hline 12 & 3 & -1 & 1 & -1 & 1 & 0 & -1 & 1 & 0 & 0 \\ 13 & -1 & 3 & -1 & -1 & 0 & 1 & -1 & 0 & 1 & 0 \\ 23 & 1 & -1 & 3 & 0 & -1 & 1 & 0 & -1 & 1 & 0 \\ 14 & -1 & -1 & 0 & 3 & -1 & -1 & -1 & 0 & 0 & 1 \\ 24 & 1 & 0 & -1 & -1 & 3 & -1 & 0 & -1 & 0 & 1 \\ 34 & 0 & 1 & 1 & -1 & -1 & 3 & 0 & 0 & -1 & 1 \\ 15 & -1 & -1 & 0 & -1 & 0 & 0 & 3 & -1 & -1 & -1 \\ 25 & 1 & 0 & -1 & 0 & -1 & 0 & -1 & 3 & -1 & -1 \\ 35 & 0 & 1 & 1 & 0 & 0 & -1 & -1 & -1 & 3 & -1 \\ 45 & 0 & 0 & 0 & 1 & 1 & 1 & -1 & -1 & -1 & 3 \end{array}.$$

Здесь индексы за границей матрицы D соответствуют номерам элементов исходной матрицы.

Далее получим

$$\begin{cases} a_{13}^3 a_{23}^{-1} a_{35}^1 = 4^1 \cdot 2^1 \cdot (1/2)^0 \cdot (1/5)^{-1} \cdot 6^1 \cdot (4/3)^0 \cdot 3^0 = 240, \\ a_{13}^{-1} a_{23}^3 a_{35}^1 = 4^{-1} \cdot 2^0 \cdot (1/2)^1 \cdot (1/5)^{-1} \cdot 6^0 \cdot (4/3)^1 \cdot 3^0 = 5/6, \\ a_{13}^1 a_{23}^1 a_{35}^3 = 4^0 \cdot 2^0 \cdot (1/2)^0 \cdot (1/5)^1 \cdot 6^1 \cdot (4/3)^1 \cdot 3^1 = 24/5; \end{cases}$$

$$P = \begin{vmatrix} 3 & -1 & 1 \\ -1 & 3 & 1 \\ 1 & 1 & 3 \end{vmatrix}; \quad P^{-1} = \begin{vmatrix} 1/2 & 1/4 & -1/4 \\ 1/4 & 1/2 & -1/4 \\ -1/4 & -1/4 & 1/2 \end{vmatrix};$$

$$\begin{vmatrix} a_{13} \\ a_{23} \\ a_{35} \end{vmatrix} = \begin{vmatrix} 240^{1/2} \cdot (5/6)^{1/4} \cdot (24/5)^{-1/4} \\ 240^{1/4} \cdot (5/6)^{1/2} \cdot (24/5)^{-1/4} \\ 240^{-1/4} \cdot (5/6)^{-1/4} \cdot (24/5)^{1/2} \end{vmatrix} = \begin{vmatrix} 10 \\ 5/\sqrt[4]{18} \\ \sqrt[4]{72}/5 \end{vmatrix} \approx \begin{vmatrix} 10 \\ 2,43 \\ 0,58 \end{vmatrix}.$$

Абсолютное значение показателя согласованности полученной матрицы, рассчитанное по формуле (4), равно 0,035. Для сравнения показателей согласованности матриц удобнее использовать не абсолютные значения, а относительные, нормированные в диапазоне $[0, 1]$. Нормировать показатель согласованности будем с учётом порядка матрицы N и максимально допустимой кратности предпочтений a_{\max} . В работе [7] предложено рассчитывать нормированный показатель согласованности по формуле

$$C = 1 - \frac{s}{s_{\max}(a_{\max}, N)},$$

где $s_{\max}(a_{\max}, N) = \begin{cases} \frac{N^3 - N^2}{2} \cdot \ln^2(a_{\max}), & \text{если } N \text{ нечетно,} \\ \left(\frac{N^3 - N^2}{2} - N\right) \cdot \ln^2(a_{\max}), & \text{если } N \text{ четно.} \end{cases}$

Пусть в нашем примере $a_{\max} = 10$. Тогда $s_{\max}(10, 5) = 265,09$; $C = 0,9999$.

Для сравнения рассчитаем индекс и коэффициент согласованности (КС) по Саати: ИС = 0,0003; КС = 0,9997.

Полученная матрица имеет высокий показатель согласованности. Согласованность не идеальна ввиду изначально содержавшейся в матрице несогласованности.

В рассмотренном примере оптимизируемой матрице соответствует связный граф. Для матрицы парных сравнений, которой соответствует несвязный граф, будет получено семейство решений.

Результаты доопределения зависят от выбранной целевой функции (функции согласованности). Результаты расчета по формуле (7) с использованием целевой функции согласованности (4) могут отличаться от результатов, полученных при оптимизации с использованием собственного числа (см. формулу (2) и работу [5]), в частности при наличии в графе доопределяемой матрицы хотя бы двух несогласованных не имеющих общих ребер троек предпочтений. Различие элементов доопределенных разными способами матриц в серии проведенных экспериментов не превышало 5 %.

Определим вычислительную сложность полученного алгоритма. Формулу (7) перепишем в виде

$$a_{i_l j_l} = \prod_{k=1}^M (q_k)^{(P^{-1})_{lk}}, \quad q_l = \prod_{\substack{(i,j) \neq (i_t, j_t), \\ i < j, t=1, \dots, M}} a_{ij}^{-d_{r(i_l, j_l) r(i, j)}}, \quad l = 1, \dots, M.$$

Сложность построения матрицы P равна $O(M^2)$; вычисления $P^{-1} - O(M^3)$. Для вычисления каждого q_l используется произведение не больше чем $N(N - 1)/2$ элементов, и количество q_l равно M . Сложность построения вектора решений по известным P^{-1} и $q_l - O(M^2)$, так как его размер равен M и каждый элемент равен произведению M сомножителей. Тогда общая сложность алгоритма равна $O(M^2) + O(M^3) + O(M \cdot N^2) + O(M^2) = O(M^3 + M \cdot N^2)$. С учетом того, что число неизвестных элементов матрицы M не может превышать общего числа пар элементов матрицы $N(N - 1)/2$, получим оценку вычислительной сложности сверху $O(N^6)$.

Заключение

Предложен способ доопределения матриц парных сравнений кратности предпочтений, вычислительная сложность которого равна $O(N^6)$, где N — порядок матрицы. Для выбранной функции оценки согласованности доопределенная матрица имеет максимально возможное значение показателя согласованности.

ЛИТЕРАТУРА

1. Саати Т., Кернс К. Аналитическое планирование. Организация систем. М.: Радио и связь, 1991.
2. Ногин В. Д. Принятие решений в многокритериальной среде: количественный подход. М.: Физматлит, 2004. 176 с.
3. Ногин В. Д. Упрощенный вариант метода анализа иерархий на основе нелинейной свертки критериев // ЖВМиМФ. 2004. Т. 44. № 7. С. 1259–1268.
4. Тоценко В. Г. Методы и системы поддержки принятия решений. Киев: Наукова думка, 2002. 381 с.
5. Микони С. В., Киселев И. С. Приближенный метод доопределения матрицы парных сравнений с кратными предпочтениями // Труды конф. IEEE AIS'07 и CAD-2007, Дивноморское, 3–10.09.2007. М.: Наука, Физматлит, 2007. С. 330–333.
6. Микони С. В., Киселев И. С. Экспериментальное доказательство достоверности метода доопределения матриц предпочтений // Материалы IV науч.-практич. конф. ИММОД-2009, 21–23.10.2009. СПб.: ФГУП ЦНИИТС, 2009. Т. 2. С. 109–112.
7. Киселев И. С. Показатель согласованности количественных предпочтений в матрице парных сравнений // Изв. Томского политехнического университета. 2011. № 5. С. 22–24.

СВЕДЕНИЯ ОБ АВТОРАХ

АЛЕКСЕЙЧУК Антон Николаевич — доктор технических наук, профессор кафедры Института специальной связи и защиты информации Национального технического университета Украины «Киевский политехнический институт», г. Киев.

E-mail: alex-crypto@mail.ru

БЫКОВА Валентина Владимировна — доцент, кандидат технических наук, профессор Института математики Сибирского федерального университета, г. Красноярск.

E-mail: bykvalen@mail.ru

ВОРОНЕНКО Андрей Анатольевич — доктор физико-математических наук, профессор кафедры математической кибернетики Московского государственного университета им. М. В. Ломоносова, г. Москва. E-mail: dm6@cs.msu.ru

ГРЕЧНИКОВ Евгений Александрович — аспирант механико-математического факультета Московского государственного университета им. М. В. Ломоносова, г. Москва. E-mail: grechnik@mccme.ru

ИЛЬЕВ Виктор Петрович — доцент, доктор физико-математических наук, доцент кафедры прикладной и вычислительной математики Омского государственного университета им. Ф. М. Достоевского, г. Омск. E-mail: iljev@mail.ru

КИСЕЛЕВ Игорь Сергеевич — аспирант кафедры математики и моделирования Петербургского государственного университета путей сообщения, г. Санкт-Петербург. E-mail: igor@kiselev.spb.ru

МАГОМЕДОВ Абдулкарим Магомедович — доцент, кандидат физико-математических наук, заведующий кафедрой дискретной математики и информатики Дагестанского государственного университета, г. Махачкала.

E-mail: magomedtagir1@yandex.ru

МАГОМЕДОВ Тагир Абдулкаримович — аспирант кафедры дискретной математики и информатики Дагестанского государственного университета, г. Махачкала.

E-mail: magomedtagir1@yandex.ru

МОНАХОВА Эмилия Анатольевна — доцент, кандидат технических наук, старший научный сотрудник Института вычислительной математики и математической геофизики СО РАН, г. Новосибирск. E-mail: emilia@rav.sccc.ru

НАВРОЦКАЯ Анна Александровна — ассистент кафедры алгебры Омского государственного университета им. Ф. М. Достоевского, г. Омск. E-mail: nawrocki@ya.ru

СТЕФАНЦОВ Дмитрий Александрович — аспирант Томского государственного университета, г. Томск. E-mail: d.a.stefantsov@isc.tsu.ru

ФОМИЧЁВ Владимир Михайлович — старший научный сотрудник, доцент, доктор физико-математических наук, ведущий научный сотрудник Учреждения Российской академии наук «Институт проблем информатики РАН», г. Москва.

E-mail: fomichev@nm.ru

ШЕВЦОВ Артур Сергеевич — преподаватель кафедры Института специальной связи и защиты информации Национального технического университета Украины «Киевский политехнический институт», г. Киев. E-mail: ashef@mail.ru

АННОТАЦИИ СТАТЕЙ НА АНГЛИЙСКОМ ЯЗЫКЕ

Alekseychuk A. N., Shevtsov A. S. **FAST ALGORITHM FOR STATISTICAL ESTIMATION OF THE MAXIMAL IMBALANCE OF BILINEAR APPROXIMATIONS OF BOOLEAN MAPPINGS.** We propose a probabilistic algorithm for determining the upper bounds of the maximal imbalance (in a given class) of bilinear approximations of Boolean mappings of n variables for a time linearly dependent on n .

Keywords: *block cipher, bilinear cryptanalysis, Boolean mapping, bilinear approximation, probabilistic algorithm.*

Voronenko A. A. **ON THE COMPLEXITY OF PROVING THAT A BOOLEAN FUNCTION IS NOT A BINARY READ-ONCE.** We show that it is sufficiently to present a linear number of values of a given Boolean function to prove that it is not read-once over the binary basis.

Keywords: *read-once Boolean function, proof complexity.*

Grechnikov E. A. **METHOD FOR CONSTRUCTING ELLIPTIC CURVES USING COMPLEX MULTIPLICATION AND ITS OPTIMIZATIONS.** Elliptic curves over finite fields with predefined conditions on the order are practically constructed using the theory of complex multiplication. A stage with the longest calculations in this method reconstructs some polynomial with integer coefficients. We prove some theoretical results and give a detailed account of the method itself and show how one can use a divisor of the mentioned polynomial with coefficients in an extension of the rational number field.

Keywords: *elliptic curves, finite fields, complex multiplication, simultaneous approximations.*

Stefantsov D. A. **IMPLEMENTATION OF SECURITY POLICIES IN PROGRAMMING INFORMATION PROCESSING SYSTEMS.** The problem of protecting information processing systems by implementing security policies in them is considered. The existing methods for solving this problem are analyzed, their disadvantages are noted, and the original method is proposed which avoids the noted disadvantages and is based on the aspect-oriented programming. In contrast to traditional aspect-oriented programming implementations, in the proposed method the security policy aspect is joined to the information processing system with the special integration module and without modification of either the information processing system or the security policy aspect that are written independently from each other and from the integration module. For the implementation of the method, the instrumental environment is created including the aspect-oriented programming language AspectTalk, the virtual machine and the translator from AspectTalk into the virtual machine language. The article contains the brief description of both the proposed method and the noted instrumental environment.

Keywords: *information processing systems, security policy, aspect-oriented programming, AspectTalk, virtual machine.*

Bykova V. V. **COMPUTATIONAL ASPECTS OF GRAPH TREewidth.** The paper gives a brief overview of recent results on the graph treewidth problem. We investigate some of the lower and upper bounds for treewidth, and present algorithmic methods to

improve these bounds.

Keywords: *graph algorithms, treewidth, partial k -tree.*

Il'ev V. P., Navrocka A. A. **COMPUTATIONAL COMPLEXITY OF THE PROBLEM OF APPROXIMATION BY GRAPHS WITH CONNECTED COMPONENTS OF BOUNDED SIZE.** New versions of the graph approximation problem with the bounded size of connected components in approximating graphs are proposed. It is shown that if the cardinality of each component in the approximating graphs is less or equal to the given integer $p \geq 3$ then the graph approximation problem is *NP*-hard, whereas in the case of $p = 2$ the problem is solvable in a polynomial time.

Keywords: *graph approximation, polynomial-time problem, NP-hard problem.*

Magomedov A. M., Magomedov T. A. **REGULAR EDGE 5-COLORING OF BIPARTITE GRAPH THAT IS AN INTERVAL ON ONE PART.** For a bipartite graph $G = (X, Y, E)$ where the degree of any vertex in X equals 2 and maximal degree of the vertex in Y equals 5, conditions for existence of regular edge 5-coloring being an interval on X are found.

Keywords: *bipartite graph, edge-coloring, NP-completeness.*

Monakhova E. A. **STRUCTURAL AND COMMUNICATIVE PROPERTIES OF CIRCULANT NETWORKS.** Circulant graphs have been extensively investigated over the past 30 years and have the broad application to different fields of computer science and discrete mathematics. Two surveys on circulant networks have been published in English: by Bermond, Comellas and Hsu (1995) and by Hwang (2003). In Russian, a survey on circulant networks is presented in a book of Monakhov and Monakhova (2000). The present paper includes the results which have not been presented in these works, and also some new results in the area of undirected circulant networks research obtained during the last years.

Keywords: *interconnection networks, circulant graphs, diameter, routing, broadcasting and gossiping.*

Fomichev V. M. **ON IMPLEMENTATION OF THE MEET-IN-THE-MIDDLE ATTACK BY MEANS OF PARALLEL COMPUTATIONS.** Three variants of implementation of the meet-in-the-middle attack based on clusters and distributed computations are considered for symmetric block cryptosystems. The average time of computations is estimated on the universal supposition of cryptosystem key equiprobability. It is shown that the reduction of the calculation time is proportional to the number of processors in the system.

Keywords: *meet-in-the-middle attack, cluster computations, distributed computations.*

Kiselev I. S. **ANALYTICAL METHOD FOR MAKING DEFINITE MULTIPLE PREFERENCES IN THE MATRIX OF INCOMPLETE PAIRWISE COMPARISONS.** The method named in the title is proposed. It has a polynomial complexity both on the size and on the number of undefined elements of the matrix, and does not make worse the consistency measure of the latter.

Keywords: *pairwise comparison matrices, consistency, incomplete pairwise comparisons.*

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, а также в перечень журналов, рекомендованных УМО в области информационной безопасности РФ 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 vestnik.tsu.ru/pdm и на Общероссийском математическом портале www.mathnet.ru. На сайте журнала можно найти также и 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 надежности 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 ее приложениям*