МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ

# ПРИКЛАДНАЯ ДИСКРЕТНАЯ МАТЕМАТИКА

*Научный журнал*

2020 № 49

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

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

Подписной индекс в объединённом каталоге «Пресса России» 38696

**УЧРЕДИТЕЛЬ**
**Томский государственный университет**



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

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

Периодичность выхода журнала: 4 номера в год.

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

Подписано к печати 03.09.2020. Формат $60 \times 84\frac{1}{8}$. Усл. п. л. 14,8. Тираж 300 экз.
Заказ № 4395. Цена свободная. Дата выхода в свет 11.09.2020.

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

# СОДЕРЖАНИЕ


## ТЕОРЕТИЧЕСКИЕ ОСНОВЫ ПРИКЛАДНОЙ ДИСКРЕТНОЙ МАТЕМАТИКИ

**Миронкин В. О.** Об образах и прообразах в графе композиции независимых равновероятных случайных отображений ..... 5

**Kutsenko A. V., Tokareva N. N.** Metrical properties of the set of bent functions in view of duality ..... 18

**Oblaukhov A. K.** On metric complements and metric regularity in finite metric spaces ..... 35

## МАТЕМАТИЧЕСКИЕ МЕТОДЫ КРИПТОГРАФИИ

**Перов А. А., Пестунов А. И.** О возможности применения свёрточных нейронных сетей к построению универсальных атак на итеративные блочные шифры ..... 46

**Черемушкин А. В.** Оценка вероятности выигрыша при проведении майнинга небольшой группой участников ..... 57

## МАТЕМАТИЧЕСКИЕ ОСНОВЫ КОМПЬЮТЕРНОЙ БЕЗОПАСНОСТИ

**Алексеев Е. К., Ахметзянова Л. Р., Бабуева А. А., Смышляев С. В.** Защищённое хранение данных и полнодисковое шифрование ..... 78

## МАТЕМАТИЧЕСКИЕ ОСНОВЫ НАДЁЖНОСТИ ВЫЧИСЛИТЕЛЬНЫХ И УПРАВЛЯЮЩИХ СИСТЕМ

**Алехина М. А.** О надёжности схем во всех полных базисах из трёхвходовых элементов при неисправностях типа 0 на выходах элементов ..... 98

## ПРИКЛАДНАЯ ТЕОРИЯ ГРАФОВ

**Монахова Э. А.** Параметрическое задание серии семейств аналитически описываемых циркулянтных сетей степени шесть ..... 108

## МАТЕМАТИЧЕСКИЕ ОСНОВЫ ИНФОРМАТИКИ И ПРОГРАММИРОВАНИЯ

**Рыбалов А. Н.** О генерической сложности экзистенциальных теорий ..... 120

СВЕДЕНИЯ ОБ АВТОРАХ ..... 127

# CONTENTS


## THEORETICAL BACKGROUNDS OF APPLIED DISCRETE MATHEMATICS

**Mironkin V. O.** On images and pre-images in a graph of the composition of independent uniform random mappings ..... 5

**Kutsenko A. V., Tokareva N. N.** Metrical properties of the set of bent functions in view of duality ..... 18

**Oblaukhov A. K.** On metric complements and metric regularity in finite metric spaces ..... 35

## MATHEMATICAL METHODS OF CRYPTOGRAPHY

**Perov A. A., Pestunov A. I.** On possibility of using convolutional neural networks for creating universal attacks on iterative block ciphers ..... 46

**Cheremushkin A. V.** Selfish mining strategy elaboration ..... 57

## MATHEMATICAL BACKGROUNDS OF COMPUTER SECURITY

**Alekseev E. K., Akhmetzyanova L. R., Babueva A. A., Smyshlyaev S. V.** Data storage security and full disk encryption ..... 78

## MATHEMATICAL BACKGROUNDS OF COMPUTER AND CONTROL SYSTEM RELIABILITY

**Alekhina M. A.** About the reliability of logic circuits in all complete bases with three-input elements and failures of zero type on their outputs ..... 98

## APPLIED GRAPH THEORY

**Monakhova E. A.** A set of families of analytically described triple loop networks defined by a parameter ..... 108

## MATHEMATICAL BACKGROUNDS OF INFORMATICS AND PROGRAMMING

**Rybalov A. N.** On generic complexity of the existential theories ..... 120

BRIEF INFORMATION ABOUT THE AUTHORS ..... 127

# ТЕОРЕТИЧЕСКИЕ ОСНОВЫ ПРИКЛАДНОЙ ДИСКРЕТНОЙ МАТЕМАТИКИ

УДК 519.212.2+519.719.2

## ОБ ОБРАЗАХ И ПРООБРАЗАХ В ГРАФЕ КОМПОЗИЦИИ НЕЗАВИСИМЫХ РАВНОВЕРОЯТНЫХ СЛУЧАЙНЫХ ОТОБРАЖЕНИЙ

В. О. Миронкин

*Национальный исследовательский университет «Высшая школа экономики», г. Москва, Россия*

Изучаются вероятностные характеристики графа случайного отображения $f_{[k]}$ — композиции $k$ независимых равновероятных случайных отображений $f_1, \ldots, f_k$, где $f_i\colon \{1, \ldots, n\} \to \{1, \ldots, n\}$, $n, k \in \mathbb{N}$, $i = 1, \ldots, n$. Получены формулы для распределения длины отрезка апериодичности произвольной вершины в графе отображения $f_{[k]}$ с учётом ряда ограничений. Выписаны формулы для вероятностей принадлежности вершины множеству $f_{[k]}(\{1, \ldots, n\})$ и множеству висячих вершин в графе отображения $f_{[k]}$. Вычислены вероятности инцидентности двух произвольных вершин одной компоненте связности, попадания произвольной вершины в множество прообразов другой вершины, а также появления коллизии в указанном графе.

**Ключевые слова:** *равновероятное случайное отображение, композиция отображений, граф отображения, образ множества, прообраз вершины, висячая вершина, слой в графе, отрезок апериодичности, коллизия.*

DOI 10.17223/20710410/49/1

## ON IMAGES AND PRE-IMAGES IN A GRAPH OF THE COMPOSITION OF INDEPENDENT UNIFORM RANDOM MAPPINGS

V. O. Mironkin

*National Research University Higher School of Economics, Moscow, Russia*

**E-mail:** mironkin.v@mail.ru

We study the probability characteristics of the random mapping graph $f_{[k]}$ — the composition of $k$ independent equiprobable random mappings $f_1, \ldots, f_k$, where $f_i\colon \{1, \ldots, n\} \to \{1, \ldots, n\}$, $n, k \in \mathbb{N}$, $i = 1, \ldots, n$. The following results are obtained. For any fixed $x, y \in S = \{1, \ldots, n\}$, $x \neq y$,

$$\mathbf{P}\{f_{[k]}(x) = f_{[k]}(y)\} = \sum_{\substack{s_1, \ldots, s_{k-1} \in \mathbb{N}\colon \\ 2 \geqslant s_1 \geqslant \ldots \geqslant s_{k-1}}} \frac{q(2, s_1)}{n^{s_{k-1}-1}} \prod_{i=1}^{k-2} q(s_i, s_{i+1}),$$

where $q(a,b) = \mathrm{C}_n^{n-b}\left(\frac{b}{n}\right)^a \sum\limits_{l=0}^{b} \mathrm{C}_b^l(-1)^l\left(1-\frac{l}{b}\right)^a$. For any fixed $x \in S$,

$$\mathbf{P}\{x \in f_{[k]}(S)\} = \frac{1}{n}\sum_{l=1}^{n}\left(\frac{(n)_l}{n^l}\right)^k +$$
$$+\sum_{l=1}^{n-2}\sum_{t=1}^{n-l-1}\sum_{m=1}^{n-t-l}(-1)^{m-1}\mathrm{C}_{n-1}^m \sum_{\substack{s_1,\ldots,s_{k-1}\in\mathbb{N}:\\ m\geqslant s_1\geqslant\ldots\geqslant s_{k-1}}} \frac{q(m,s_1)}{n^{s_{k-1}}}\prod_{i=1}^{k-2} q(s_i,s_{i+1})V_{s_1,\ldots,s_{k-1}}^{\{k,m\}},$$

where

$$V_{s_1,\ldots,s_{k-1}}^{\{k,m\}} = \mathbf{P}\{x \in H_{f_{[k]}}^{(t,l)} \mid D_{s_1,\ldots,s_{k-1},1}^{\{k\}}(y_1,\ldots,y_m), f_{[k]}(y_1) = x\} =$$
$$= \frac{1}{n}\prod_{i=m+1}^{t+l+m-1}\left(1-\frac{i}{n}\right)\prod_{i=1}^{k-1}\prod_{j=s_i+1}^{t+l+s_i-2}\left(1-\frac{j}{n}\right)\sum_{v=0}^{k-1}\prod_{u=1}^{v}\left(1-\frac{t+l+s_u-1}{n}\right),$$

$H_f^{(t,l)}$ is $t$-th layer of cycles of length $l$ in graph $G_f$, $D_{s_1,\ldots,s_k}^{\{k\}}(y_1,\ldots,y_m) = \bigcap\limits_{i=1}^{k}\{|\{f_{[i]}(y_1),\ldots,f_{[i]}(y_m)\}| = s_i\}$, and $(n)_z = n(n-1)\ldots(n-z+1)$. For any fixed $x \in S \setminus S'$ and for any $r \in \{1,\ldots,n-1\}$, $S' \subseteq S$, $|S'| = r$, $z \in \{1,\ldots,n\}$,

$$\mathbf{P}\{\tau_{f_{[k]}}(x) = z, \mathcal{R}_{f_{[k]}}(x) \cap S' = \varnothing\} =$$
$$= \left(1-\left(1-\frac{z}{n}\right)\left(1-\frac{z-1}{n}\right)^{k-1}\right)\left(\frac{(n)_{z-1}}{n^{z-1}}\right)^{k-1}\frac{(n)_{r+z}}{n^{z-1}\,(n)_{r+1}},$$

where $\mathcal{R}_{f_{[k]}}(x)$ is the aperiodicity segment of vertex $x$ in the graph of mapping $f_{[k]}$, $\tau_{f_{[k]}}(x) = \min\{t \in \mathbb{N}\colon {f_{[k]}}^t(x) \in \{x, f_{[k]}(x),\ldots,{f_{[k]}}^{t-1}(x)\}\}$. For any fixed $x, y \in S$, $x \neq y$, and for any $r \in \{1,\ldots,n\}$,

$$\mathbf{P}\{y \in (f_{[k]})^{-r}(x)\} = \frac{1}{n}\left(1-\frac{1}{n-1}\sum_{z\in Q_r\setminus\{1\}}\left(\frac{(n)_z}{n^z}\right)^k\right),$$

where $Q_r = \{m \in \mathbb{N}\colon m|r\}$.

**Keywords:** *equiprobable random mapping, composition of mappings, graph of a mapping, image of a multitude, pre-image of a vertex, initial vertex, layer in a graph, aperiodicity segment, collision.*

## Введение

Настоящая работа продолжает цикл работ [1, 2], посвящённых изучению вероятностных свойств и характеристик композиции независимых равновероятных случайных отображений [3 – 7] — математического объекта, используемого при моделировании итерационных механизмов защиты информации, в том числе алгоритмов выработки производных ключей [8, 9], итерации которых строятся с помощью разных процедур и разных случайных элементов (например, раундовых ключей, векторов инициализации).

Аналогично [1], рассмотрим конечное множество $S = \{1,\ldots,n\}$, $n > 1$, и вероятностное пространство $(\Omega, \mathcal{F}, \mathbf{P})$, в котором пространством элементарных исходов $\Omega$ является множество $\mathfrak{S}$ всех $n^n$ отображений $f\colon S \to S$, алгеброй событий $\mathcal{F}$ — множество всех подмножеств $\Omega$, а вероятностная мера $\mathbf{P}$, соответствующая равновероятным случайным отображениям, задана следующим образом:

$$\forall f \in \Omega\ \left(\mathbf{P}(f) = n^{-n}\right). \tag{1}$$

Будем использовать следующие определения для характеристик графа отображения (см. также [10 – 13]; отображение $f$ считается детерминированным).

**Определение 1.** *Графом отображения* $f$ называется ориентированный граф $G_f = (S, E_f)$ с множеством вершин $S$ и множеством ориентированных рёбер $E_f = \{(x, f(x))\colon x \in S\} \subset S^2$.

**Определение 2.** *Компонентой связности* $\mathcal{K}_f(x)$ графа $G_f$, содержащей вершину $x \in S$, называется множество вершин

$$\left\{y \in S\colon f^l(y) = f^k(x) \text{ для некоторых } k, l \geqslant 0\right\}.$$

**Определение 3.** Вершина $x \in S$ называется *циклической вершиной* графа $G_f$ отображения $f$, если существует такое $b \geqslant 1$, что $f^b(x) = x$.

Обозначим: $C(G_f)$ — множество циклических вершин графа $G_f$; $C_l(G_f)$ — множество вершин, лежащих на циклах длины $l \in \{1, \ldots, n\}$; $\beta_f(x)$ — длина цикла компоненты $\mathcal{K}_f(x)$.

**Определение 4.** *Высотой* $\alpha_f(x)$ вершины $x \in S$ в графе $G_f$ называется расстояние от этой вершины до ближайшей циклической вершины:

$$\alpha_f(x) = \min\{m \geqslant 0\colon f^m(x) \in C(G_f)\}.$$

**Определение 5.** *Отрезком апериодичности* $\mathcal{R}_f(x)$, начинающимся в вершине $x \in S$ графа $G_f$, называется отрезок выходящей из $x$ траектории от $x$ до её первого самопересечения.

Через $\tau_f(x)$ обозначим случайную величину, равную длине отрезка апериодичности $\mathcal{R}_f(x)$:

$$\tau_f(x) = \min\left\{t \in \mathbb{N}\colon f^t(x) \in \left\{x, f(x), \ldots, f^{t-1}(x)\right\}\right\}.$$

Как и в [1], зависимость случайных величин $\alpha_f(x)$, $\beta_f(x)$ и $\tau_f(x)$ от параметра $n$ отображать не будем.

**Определение 6.** Для произвольных $l \in \{1, \ldots, n\}$, $t \in \{0, \ldots, n-l\}$ назовём *$t$-м слоем циклов длины $l$* в графе $G_f$ множество вершин

$$H_f^{(t,l)} = \{x \in S\colon \alpha_f(x) = t, \beta_f(x) = l\}.$$

Далее для произвольного $k \in \mathbb{N}$ рассмотрим последовательность независимых отображений $f_1, \ldots, f_k$, имеющих распределение (1) на $\mathfrak{S}$. Через $f_{[k]}$ обозначим композицию отображений: $f_k(\ldots(f_1(x))\ldots)$, $x \in S$; $f_{[0]}$ будем понимать как тождественное отображение.

Отметим, что если случайные отображения $f_1, \ldots, f_k$ имеют равновероятное распределение (1), то распределение $f_{[k]}$ при $k > 1$ не является равновероятным на $\mathfrak{S}$, так как $\left|f_{[1]}(S)\right| \geqslant \left|f_{[2]}(S)\right| \geqslant \ldots$

В настоящей работе изучаются вероятностные характеристики множества $f_{[k]}(S)$ и множества прообразов произвольной вершины $x \in S$ в случае, когда $k$ произвольное и случайные отображения $f_1, \ldots, f_k$ независимы и имеют распределение (1) на $\mathfrak{S}$.

## 1. Образ множества $S$ и коллизии при отображении $f_{[k]}$

В рамках решения задач, связанных с изучением множества $f_{[k]}(S)$, в [4] получены оценки среднего размера образа подмножества множества $S$ при действии композиции случайных отображений.

Результаты, приведённые в данной работе, позволяют выписать точные формулы для ряда вероятностных характеристик образа $f_{[k]}(S)$ исходного множества $S$, в том числе для его среднего размера.

Для произвольных $k, m, s_1, \ldots, s_k \in \mathbb{N}$, $n \geqslant m \geqslant s_1 \geqslant \ldots \geqslant s_k$, и произвольных фиксированных различных вершин $y_1, \ldots, y_m \in S$ рассмотрим событие

$$D^{\{k\}}_{s_1,\ldots,s_k}(y_1, \ldots, y_m) = \bigcap_{i=1}^{k} \{|\{f_{[i]}(y_1), \ldots, f_{[i]}(y_m)\}| = s_i\}.$$

**Лемма 1.** Пусть $k \in \mathbb{N}$ — произвольное, случайные отображения $f_1, \ldots, f_k$ независимы и имеют распределение (1) на $\mathfrak{S}$. Тогда для любых $m, s_1, \ldots, s_k \in \mathbb{N}$, $n \geqslant m \geqslant s_1 \geqslant \ldots \geqslant s_k$, и любых фиксированных различных $y_1, \ldots, y_m \in S$ справедливо равенство

$$\mathbf{P}\{D^{\{k\}}_{s_1,\ldots,s_k}(y_1, \ldots, y_m)\} = q(m, s_1) \prod_{i=1}^{k-1} q(s_i, s_{i+1}), \tag{2}$$

где $q(a, b) = \mathrm{C}_n^{n-b} \left(\dfrac{b}{n}\right)^a \sum\limits_{l=0}^{b} \mathrm{C}_b^l (-1)^l \left(1 - \dfrac{l}{b}\right)^a$.

***Доказательство.*** Для произвольных $i \in \{1, \ldots, k\}$, $a \in \{1, \ldots, m\}$ и $b \in \{1, \ldots, a\}$ определим событие

$$D^{(i)}(a, b) = \{|\{f_i(1), \ldots, f_i(a)\}| = b\},$$

вероятность которого, согласно [14], равна

$$q^{(i)}(a, b) = \mathbf{P}\{D^{(i)}(a, b)\} = \mathbf{P}\{\mu_0(a, n) = n - b\} = \mathrm{C}_n^{n-b} \left(\frac{b}{n}\right)^a \sum_{l=0}^{b} \mathrm{C}_b^l (-1)^l \left(1 - \frac{l}{b}\right)^a, \tag{3}$$

где случайная величина $\mu_0(a, n)$ — число пустых ячеек в схеме равновероятных размещений, в которой $a$ частиц независимо друг от друга размещаются в $n$ ячейках [14].

Поскольку из полученного выражения следует, что $q^{(i)}(a, b)$ не зависят от $i$, верхний индекс в обозначении $q^{(i)}(a, b)$ опустим.

С учётом (3) в силу независимости отображений $f_1, \ldots, f_k$ для произвольного $m \in \mathbb{N}$, произвольных фиксированных различных вершин $y_1, \ldots, y_m \in S$ и произвольного набора $(s_1, \ldots, s_k)$, такого, что $n \geqslant m \geqslant s_1 \geqslant \ldots \geqslant s_k$, получаем равенство (2). ■

**Определение 7.** *Коллизией* в графе отображения $G_f$ называется произвольная пара вершин $x, y \in S$, $x \neq y$, для которых $f(x) = f(y)$.

Лемма 1 позволяет вычислить вероятность события, состоящего в том, что произвольные фиксированные вершины $x, y \in S$, $x \neq y$, образуют коллизию в графе отображения $G_{f_{[k]}}$.

**Теорема 1.** Пусть $k \in \mathbb{N}$ — произвольное, случайные отображения $f_1, \ldots, f_k$ независимы и имеют распределение (1) на $\mathfrak{S}$. Тогда для любых фиксированных $x, y \in S$, $x \neq y$, справедливо равенство

$$\mathbf{P}\{f_{[k]}(x) = f_{[k]}(y)\} = \sum_{\substack{s_1,\ldots,s_{k-1}: \\ 2 \geqslant s_1 \geqslant \ldots \geqslant s_{k-1}}} \frac{q(2, s_1)}{n^{s_{k-1}-1}} \prod_{i=1}^{k-2} q(s_i, s_{i+1}),$$

где $q(a, b) = \mathrm{C}_n^{n-b} \left(\dfrac{b}{n}\right)^a \sum\limits_{l=0}^{b} \mathrm{C}_b^l (-1)^l \left(1 - \dfrac{l}{b}\right)^a$.

***Доказательство.*** Для произвольных фиксированных $x, y \in S$, $x \neq y$, с учётом леммы 1 имеет место цепочка соотношений

$$\mathbf{P}\{f_{[k]}(x) = f_{[k]}(y)\} = \sum_{\substack{s_1,\ldots,s_{k-1}:\\ 2\geqslant s_1\geqslant\ldots\geqslant s_{k-1}}} \mathbf{P}\{D^{\{k\}}_{s_1,\ldots,s_{k-1},1}(x,y)\} =$$

$$= \sum_{\substack{s_1,\ldots,s_{k-1}:\\ 2\geqslant s_1\geqslant\ldots\geqslant s_{k-1}}} q(2,s_1)q(s_{k-1},1)\prod_{i=1}^{k-2} q(s_i,s_{i+1}) = \sum_{\substack{s_1,\ldots,s_{k-1}:\\ 2\geqslant s_1\geqslant\ldots\geqslant s_{k-1}}} \frac{q(2,s_1)}{n^{s_{k-1}-1}}\prod_{i=1}^{k-2} q(s_i,s_{i+1}).$$

Теорема доказана. ■

**Замечание 1.** Согласно теореме 1, в силу равноправия вершин из $S$ среднее число коллизий в графе $G_{f_{[k]}}$ определяется величиной

$$\mathrm{C}_n^2\mathbf{P}\{f_{[k]}(x) = f_{[k]}(y)\}.$$

**Определение 8.** Вершина $x \in S$ в графе $G_f$ называется *висячей*, если не существует $y \in S$, для которого $f(y) = x$.

**Теорема 2.** Пусть $k \in \mathbb{N}$ — произвольное, случайные отображения $f_1, \ldots, f_k$ независимы и имеют распределение (1) на $\mathfrak{S}$. Тогда для любого фиксированного $x \in S$ справедливо равенство

$$\mathbf{P}\{x \in f_{[k]}(S)\} = \frac{1}{n}\sum_{l=1}^{n}\left(\frac{(n)_l}{n^l}\right)^k +$$
$$+\sum_{l=1}^{n-2}\sum_{t=1}^{n-l-1}\sum_{m=1}^{n-t-l}(-1)^{m-1}\mathrm{C}_{n-1}^m \sum_{\substack{s_1,\ldots,s_{k-1}:\\ m\geqslant s_1\geqslant\ldots\geqslant s_{k-1}}} \frac{q(m,s_1)}{n^{s_{k-1}}}\prod_{i=1}^{k-2} q(s_i,s_{i+1})V^{\{k,m\}}_{s_1,\ldots,s_{k-1}},$$

где $q(a,b) = \mathrm{C}_n^{n-b}\left(\dfrac{b}{n}\right)^a \displaystyle\sum_{l=0}^{b}\mathrm{C}_b^l(-1)^l\left(1-\frac{l}{b}\right)^a$; $V^{\{k,m\}}_{s_1,\ldots,s_{k-1}}$ определяется соотношением (9) (см. далее); $(n)_z = n(n-1)\ldots(n-z+1)$ — $z$-я факториальная степень числа $n$.

***Доказательство.*** Заметим, что произвольная фиксированная вершина $x \in S$ лежит в множестве $f_{[k]}(S)$ в случаях, когда она либо является циклической в графе $G_{f_{[k]}}$, либо лежит на подходах к циклу и при этом не является висячей. Таким образом, выполняется равенство

$$\{x \in f_{[k]}(S)\} = \{x \in C(G_{f_{[k]}})\} \cup \bigcup_{l=1}^{n-2}\bigcup_{t=1}^{n-l-1}\{x \in H^{(t,l)}_{f_{[k]}}, |(f_{[k]})^{-1}(x)| \geqslant 1\},$$

где под знаком объединения стоят несовместные события. Тогда, переходя к вероятностям событий, получаем

$$\mathbf{P}\{x \in f_{[k]}(S)\} = \mathbf{P}\{x \in C(G_{f_{[k]}})\} + \sum_{l=1}^{n-2}\sum_{t=1}^{n-l-1}\mathbf{P}\{x \in H^{(t,l)}_{f_{[k]}}, |(f_{[k]})^{-1}(x)| \geqslant 1\}. \qquad (4)$$

Согласно [2], первое слагаемое в правой части (4) равно

$$\mathbf{P}\{x \in C(G_{f_{[k]}})\} = \frac{1}{n}\sum_{l=1}^{n}\left(\frac{(n)_l}{n^l}\right)^k. \qquad (5)$$

Рассмотрим отдельно величины, стоящие под знаками суммирования в (4), при фиксированных $l \in \{1,\dots,n-2\}$, $t \in \{1,\dots,n-l-1\}$. По формуле включения-исключения в силу равноправия всех вершин $y \in S \setminus \{x\}$ имеем

$$\begin{gathered}\mathbf{P}\{x \in H^{(t,l)}_{f_{[k]}}, |(f_{[k]})^{-1}(x)| \geqslant 1\} = \mathbf{P}\{\bigcup_{y \in S\setminus\{x\}} \{x \in H^{(t,l)}_{f_{[k]}}, y \in (f_{[k]})^{-1}(x)\}\} = \\ = \sum_{m=1}^{n-t-l} (-1)^{m-1} \mathrm{C}^m_{n-1} \mathbf{P}\left\{\begin{array}{c} x \in H^{(t,l)}_{f_{[k]}}, f_{[k]}(y_1) = \ldots = f_{[k]}(y_m) = x, \\ y_1,\dots,y_m \in S\setminus\{x\} \text{ — различны} \end{array}\right\}.\end{gathered} \tag{6}$$

Рассмотрим вероятность, стоящую под знаком суммирования в правой части (6), при фиксированном значении $m \in \{1,\dots,n-l-t\}$. Для произвольных фиксированных различных $y_1,\dots,y_m \in S \setminus \{x\}$ по формуле полной вероятности имеем

$$\begin{gathered}\mathbf{P}\{x \in H^{(t,l)}_{f_{[k]}}, f_{[k]}(y_1) = \ldots = f_{[k]}(y_m) = x\} = \\ = \sum_{\substack{s_1,\dots,s_{k-1}: \\ m \geqslant s_1 \geqslant \ldots \geqslant s_{k-1}}} \mathbf{P}\{D^{\{k\}}_{s_1,\dots,s_{k-1},1}(y_1,\dots,y_m), f_{[k]}(y_1) = x, x \in H^{(t,l)}_{f_{[k]}}\} = \\ = \sum_{\substack{s_1,\dots,s_{k-1}: \\ m \geqslant s_1 \geqslant \ldots \geqslant s_{k-1}}} \mathbf{P}\{x \in H^{(t,l)}_{f_{[k]}} \mid D^{\{k\}}_{s_1,\dots,s_{k-1},1}(y_1,\dots,y_m), f_{[k]}(y_1) = x\} \times \\ \times \mathbf{P}\{D^{\{k\}}_{s_1,\dots,s_{k-1},1}(y_1,\dots,y_m), f_{[k]}(y_1) = x\}.\end{gathered} \tag{7}$$

При этом в силу независимости отображений $f_1,\dots,f_k$ с учётом леммы 1

$$\begin{gathered}\mathbf{P}\{D^{\{k\}}_{s_1,\dots,s_{k-1},1}(y_1,\dots,y_m), f_{[k]}(y_1) = x\} = \\ = \frac{1}{n^{s_{k-1}}} \mathbf{P}\{D^{\{k-1\}}_{s_1,\dots,s_{k-1}}(y_1,\dots,y_m)\} = \frac{q(m,s_1)}{n^{s_{k-1}}} \prod_{i=1}^{k-2} q(s_i, s_{i+1}).\end{gathered} \tag{8}$$

Далее заметим, что вычисление условной вероятности, стоящей под знаком суммирования в правой части (7), проводится аналогично [2] с поправкой на наличие дополнительных $s_1,\dots,s_{k-1},m$ вершин в множествах $f_{[1]}(S),\dots,f_{[k]}(S)$ соответственно, а именно:

$$\begin{gathered}V^{\{k,m\}}_{s_1,\dots,s_{k-1}} = \mathbf{P}\{x \in H^{(t,l)}_{f_{[k]}} \mid D^{\{k\}}_{s_1,\dots,s_{k-1},1}(y_1,\dots,y_m), f_{[k]}(y_1) = x\} = \\ = \frac{1}{n} \prod_{i=m+1}^{t+l+m-1} \left(1 - \frac{i}{n}\right) \prod_{i=1}^{k-1} \prod_{j=s_i+1}^{t+l+s_i-2} \left(1 - \frac{j}{n}\right) \sum_{v=0}^{k-1} \prod_{u=1}^{v} \left(1 - \frac{t+l+s_u-1}{n}\right).\end{gathered} \tag{9}$$

Подставив (8) и (9) в (7), с учётом (4)–(6) получим искомый результат. ■

**Следствие 1.** Пусть $k \in \mathbb{N}$ — произвольное, случайные отображения $f_1,\dots,f_k$ независимы и имеют распределение (1) на $\mathfrak{S}$. Тогда

$$\mathbf{E}\left|f_{[k]}(S)\right| = n\mathbf{P}\{x \in f_{[k]}(S)\}.$$

Из определения 8 следует, что множество висячих вершин графа $G_f$ совпадает с множеством вершин, не имеющих прообразов.

Через $T_{f_{[k]}}$ обозначим множество висячих вершин в графе $G_{f_{[k]}}$, $k \in \mathbb{N}$. Множество $T_f$ исследовано, например, в [13]. Рассмотрим случай $k \geqslant 2$. Найдём вероятность попадания случайной вершины графа $G_{f_{[k]}}$ в множество $T_{f_{[k]}}$.

**Замечание 2.** Из равенства $S = T_{f_{[k]}} \cup f_{[k]}(S)$, где $T_{f_{[k]}} \cap f_{[k]}(S) = \varnothing$, вытекает выражение для вероятности попадания произвольной вершины $x$ в множество висячих вершин в графе $G_{f_{[k]}}$:

$$\mathbf{P}\{x \in T_{f_{[k]}}\} = 1 - \mathbf{P}\{x \in f_{[k]}(S)\}.$$

При этом в силу равноправия вершин из $S$ выполняются соотношения

$$\mathbf{E}\left|T_{f_{[k]}}\right| = n\mathbf{P}\{x \in T_{f_{[k]}}\} = n - n\mathbf{P}\{x \in f_{[k]}(S)\}.$$

## 2. Инцидентность вершин одной компоненте связности в графе отображения $f_{[k]}$

Для вычисления вероятности попадания вершин из $S$ в одну компоненту связности графа $G_{f_{[k]}}$ докажем ряд вспомогательных утверждений.

Выделим в исходном множестве $S$ некоторое подмножество вершин $S' = \{y_1, \ldots, y_r\} \subseteq S$, где $r < n$. Для данного множества вычислим вероятность события, заключающегося в том, что отрезок апериодичности $\mathcal{R}_{f_{[k]}}(x)$ произвольной вершины $x \in S \setminus S'$ не проходит через вершины множества $S'$.

**Теорема 3.** Пусть $k \in \mathbb{N}$ — произвольное, случайные отображения $f_1, \ldots, f_k$ независимы и имеют распределение (1) на $\mathfrak{S}$. Тогда для любого фиксированного $x \in S \setminus S'$ и любых $r \in \{1, \ldots, n-1\}$, $S' \subseteq S$ ($|S'| = r$) и $z \in \{1, \ldots, n\}$ справедливо равенство

$$\mathbf{P}\left\{\tau_{f_{[k]}}(x) = z, \mathcal{R}_{f_{[k]}}(x) \cap S' = \varnothing\right\} =$$
$$= \left(1 - \left(1 - \frac{z}{n}\right)\left(1 - \frac{z-1}{n}\right)^{k-1}\right)\left(\frac{(n)_{z-1}}{n^{z-1}}\right)^{k-1} \frac{(n)_{r+z}}{n^{z-1}\,(n)_{r+1}}.$$

***Доказательство.*** Зафиксируем $S' = \{y_1, \ldots, y_r\} \subseteq S$ и $x \in S \setminus S'$. По формуле условной вероятности

$$\mathbf{P}\{\tau_{f_{[k]}}(x) = z, \mathcal{R}_{f_{[k]}}(x) \cap S' = \varnothing\} = \mathbf{P}\{\tau_{f_{[k]}}(x) = z\}\mathbf{P}\{\mathcal{R}_{f_{[k]}}(x) \cap S' = \varnothing \mid \tau_{f_{[k]}}(x) = z\} =$$
$$= \left(F_{[k]}(z) - F_{[k]}(z-1)\right)\frac{(n-r-1)_{z-1}z}{(n-1)_{z-1}z},$$

где $(n-1)_{z-1}\,z$ — общее число отрезков апериодичности длины $z$ графа $G_{f_{[k]}}$, начинающихся в вершине $x \in S$. Преобразовав полученное выражение с учётом равенства [1]

$$F_k(z) = 1 - \left(1 - \frac{z}{n}\right)\left(\frac{(n)_z}{n^z}\right)^k,$$

получим искомое равенство. ■

**Следствие 2.** Пусть $k \in \mathbb{N}$ — произвольное, случайные отображения $f_1, \ldots, f_k$ независимы и имеют распределение (1) на $\mathfrak{S}$. Тогда для любого фиксированного $x \in S \setminus S'$ и любых $r \in \{1, \ldots, n-1\}$, $S' \subseteq S$, $|S'| = r$, справедливо равенство

$$\mathbf{P}\{\mathcal{R}_{f_{[k]}}(x) \cap S' = \varnothing\} =$$
$$= \sum_{z=1}^{n-r-1}\left(1 - \left(1 - \frac{z+1}{n}\right)\left(1 - \frac{z}{n}\right)^{k-1}\right)\left(\frac{(n)_z}{n^z}\right)^{k-1}\frac{(n)_{r+z+1}}{n^z(n)_{r+1}}. \tag{10}$$

Формулы, полученные в теореме 3 и следствии 2, могут быть использованы при решении задачи оценки допустимого периода эксплуатации долговременных ключей в процессе функционирования итерационных алгоритмов типа [8] при наличии информации о «слабых» ключах.

Далее для случая $r = 1$ вычислим вероятность того, что вершина $y = y_1$, соответствующая некоторому «слабому» ключу, попадёт в компоненту связности $\mathcal{K}_{f_{[k]}}(x)$, где $x \in S \setminus \{y\}$ — произвольный фиксированный.

**Теорема 4.** Пусть $k \in \mathbb{N}$ — произвольное, случайные отображения $f_1, \ldots, f_k$ независимы и имеют распределение (1) на $\mathfrak{S}$. Тогда для любых фиксированных $x, y \in S$, $x \neq y$, справедливо равенство

$$\begin{gathered}\mathbf{P}\{y \in \mathcal{K}_{f_{[k]}}(x)\} = 1 - \sum_{z=1}^{n-2} \left(1 - \left(1 - \frac{z+1}{n}\right)\left(1 - \frac{z}{n}\right)^{k-1}\right) \left(\frac{(n)_z}{n^z}\right)^k \frac{\mathrm{C}_{n-z}^2}{\mathrm{C}_n^2} + \\ + \sum_{z=1}^{n-1} \sum_{u=1}^{n-z} \frac{z^2}{n-1} \left(\frac{(n)_{z+u}}{n^{z+u}}\right)^k + \sum_{z=1}^{n-1} \sum_{u=0}^{n-z-1} \sum_{s=1}^{k-1} \sum_{t=1}^{k-1} \chi(z,u,s,t) + \\ + \sum_{z=1}^{n-1} \sum_{u=z}^{n-1} \frac{(z^2-z)(n-u)}{u(n-1)} \left(\frac{(n)_u}{n^u}\right)^k \left(1 - \left(1 - \frac{u}{n}\right)^{k-1}\right),\end{gathered}$$

где $\chi(z,u,s,t)$ определяется соотношением (16) (см. далее).

***Доказательство.*** Для произвольной фиксированной пары вершин $x, y \in S$, $x \neq y$, выполняется равенство событий

$$\{y \in \mathcal{K}_{f_{[k]}}(x)\} = \{y \in \mathcal{R}_{f_{[k]}}(x)\} \cup \bigcup_{z=1}^{n-1} \bigcup_{u=0}^{n-z-1} \{\tau_{f_{[k]}}(x) = z, f_{[k]}^u(y) \notin \mathcal{R}_{f_{[k]}}(x), f_{[k]}^{u+1}(y) \in \mathcal{R}_{f_{[k]}}(x)\},$$

где под знаками объединения стоят несовместные события. Поэтому, переходя к вероятностям с учётом равноправия всех вершин из $S$, получаем

$$\begin{gathered}\mathbf{P}\{y \in \mathcal{K}_{f_{[k]}}(x)\} = \mathbf{P}\{y \in \mathcal{R}_{f_{[k]}}(x)\} + \\ + \sum_{z=1}^{n-1} \sum_{u=0}^{n-z-1} \mathbf{P}\{\tau_{f_{[k]}}(x) = z, f_{[k]}^u(y) \notin \mathcal{R}_{f_{[k]}}(x), f_{[k]}^{u+1}(y) \in \mathcal{R}_{f_{[k]}}(x)\}.\end{gathered} \tag{11}$$

Выражение для первого слагаемого в (11) следует из (10) при $r = 1$:

$$\begin{gathered}\mathbf{P}\{y \in \mathcal{R}_{f_{[k]}}(x)\} = 1 - \mathbf{P}\{y \notin \mathcal{R}_{f_{[k]}}(x)\} = \\ = 1 - \sum_{z=1}^{n-2} \left(1 - \left(1 - \frac{z+1}{n}\right)\left(1 - \frac{z}{n}\right)^{k-1}\right) \left(\frac{(n)_z}{n^z}\right)^k \frac{\mathrm{C}_{n-z}^2}{\mathrm{C}_n^2}.\end{gathered} \tag{12}$$

Вычислим вероятность события, стоящего под знаками суммирования в (11) при фиксированных значениях $z \in \{1, \ldots, n-1\}$ и $u \in \{0, \ldots, n-z-1\}$. Для указанных значений $z$, $u$ определим события

$$A_{z,s}^{\{x\}} = \left\{ \min_{j \in \{1,\ldots,k\}} \{j \colon f_{[j]}(f_{[k]}^{z-1}(x)) \in \{f_{[j]}(f_{[k]}^m(x)), m = 0, \ldots, z-2\}\} = s \right\},$$

$$B_{u,s}^{\{x,y\}} = \left\{ \min_{j \in \{1,\ldots,k\}} \{j \colon f_{[j]}(f_{[k]}^u(y)) \in \{f_{[i]}(f_{[k]}^m(x)), m = 0, \ldots, z-1, i = 1, \ldots, k\}\} = s \right\}.$$

Тогда выполняется равенство

$$\{\tau_{f_{[k]}}(x)=z, f^u_{[k]}(y)\notin\mathcal{R}_{f_{[k]}}(x), f^{u+1}_{[k]}(y)\in\mathcal{R}_{f_{[k]}}(x)\}=$$
$$=\bigcup_{s=1}^{k}\bigcup_{t=1}^{k}\left\{\begin{array}{c}\tau_{f_{[k]}}(x)=z, A^{\{x\}}_{z,s}, B^{\{x,y\}}_{u,t},\\ f^u_{[k]}(y)\notin\mathcal{R}_{f_{[k]}}(x), f^{u+1}_{[k]}(y)\in\mathcal{R}_{f_{[k]}}(x)\end{array}\right\}.$$

Вычислим вероятности $p_{z,u,s,t}$ событий, стоящих под знаком объединения, при фиксированных значениях $s\in\{1,\ldots,k\}$, $t\in\{1,\ldots,k\}$.

В случае $s=t=k$:

$$p_{z,u,k,k}=\frac{z^2}{n}\prod_{i=2}^{z+u}\left(1-\frac{i}{n}\right)\prod_{i=1}^{z+u}\left(1-\frac{i}{n}\right)^{k-1}=\frac{z^2}{n-1}\left(\frac{(n)_{z+u+1}}{n^{z+u+1}}\right)^k. \tag{13}$$

В случае $s=k$, $t<k$:

$$\begin{aligned}p_{z,u,k,t}&=\frac{z(z-1)}{n}\prod_{i=2}^{z+u}\left(1-\frac{i}{n}\right)\prod_{i=1}^{z+u-1}\left(1-\frac{i}{n}\right)^{k-1}\left(1-\frac{z+u}{n}\right)^{t-1}=\\&=\frac{z(z-1)}{n-1}\left(\frac{(n)_{z+u}}{n^{z+u}}\right)^k\left(1-\frac{z+u}{n}\right)^t.\end{aligned} \tag{14}$$

В случае $s<k$, $t=k$:

$$\begin{aligned}p_{z,u,s,k}&=\frac{z(z-1)}{n}\prod_{i=2}^{z+u}\left(1-\frac{i}{n}\right)\prod_{i=1}^{z+u-1}\left(1-\frac{i}{n}\right)^{k-1}\left(1-\frac{z+u}{n}\right)^{s-1}=\\&=\frac{z(z-1)}{n-1}\left(\frac{(n)_{z+u}}{n^{z+u}}\right)^k\left(1-\frac{z+u}{n}\right)^s.\end{aligned} \tag{15}$$

В случае $s<k$, $t<k$:

$$\begin{aligned}\chi(z,u,s,t)=p_{z,u,s,t}&=\frac{(z-1)^2}{n}\prod_{i=2}^{z+u}\left(1-\frac{i}{n}\right)\prod_{i=1}^{z+u-2}\left(1-\frac{i}{n}\right)^{k-1}\times\\\times\left(1-\frac{z+u-1}{n}\right)^{s-1}&\left(1-\frac{z+u}{n}\right)^{\min(t,s)-1}\left(1-\frac{z+u-1}{n}\right)^{t-\min(t,s)}=\\=\frac{(z-1)^2}{n-1}&\left(\frac{(n)_{z+u-1}}{n^{z+u-1}}\right)^k\left(1-\frac{z+u-1}{n}\right)^{s+t-\min(t,s)}\left(1-\frac{z+u}{n}\right)^{\min(t,s)}.\end{aligned} \tag{16}$$

В итоге, подставив выражения (12)–(16) в (11) и сгруппировав слагаемые, получим искомую формулу. ■

**Замечание 3.** Для средней мощности компоненты связности произвольной фиксированной вершины $x\in S$ в графе $G_{f_{[k]}}$ справедливо равенство

$$\mathbf{E}|\mathcal{K}_{f_{[k]}}(x)|=(n-1)\mathbf{P}\{y\in\mathcal{K}_{f_{[k]}}(x)\}+1.$$

### 3. Прообразы случайной вершины в графе отображения $f_{[k]}$

Для произвольного фиксированного $x\in S$ и произвольного $r\in\mathbb{N}$ через $\left(f_{[k]}\right)^{-r}(x)$ обозначим множество $\{y\in S\colon f^r_{[k]}(y)=x\}$. Дополнительно для произвольных $j,r\in\mathbb{N}$ определим

$$Q_r=\{m\in\mathbb{N}\colon m|r\}. \tag{17}$$

Заметим, что для произвольного фиксированного $x \in S$ вероятность события $\{y \in \in \left(f_{[k]}\right)^{-r}(x)\}$ зависит от выбора $y \in S$, а именно от выполнения и невыполнения условия $y = x$. Так, в частности, в случае $y = x$ выполняется равенство

$$\{x \in \left(f_{[k]}\right)^{-r}(x)\} = \bigcup_{m \in Q_r} \{x \in C_m(G_{f_{[k]}})\},$$

где под знаком объединения стоят несовместные события, и поэтому для равновероятных независимых случайных отображений $f_1, \ldots, f_k$ с учётом [1]

$$\mathbf{P}\{x \in \left(f_{[k]}\right)^{-r}(x)\} = \frac{1}{n} \sum_{m \in Q_r} \left(\frac{(n)_m}{n^m}\right)^k.$$

В случае $y \neq x$ справедлив следующий результат.

**Теорема 5.** Пусть $k \in \mathbb{N}$ — произвольное, случайные отображения $f_1, \ldots, f_k$ независимы и имеют распределение (1) на $\mathfrak{S}$. Тогда для любых фиксированных $x, y \in S$, $x \neq y$, и любого $r \in \{1, \ldots, n\}$ справедливо равенство

$$\mathbf{P}\{y \in \left(f_{[k]}\right)^{-r}(x)\} = \frac{1}{n}\left(1 - \frac{1}{n-1} \sum_{z \in Q_r \setminus \{1\}} \left(\frac{(n)_z}{n^z}\right)^k\right),$$

где $Q_r$ определяется соотношением (17).

***Доказательство.*** Для произвольных фиксированных $x, y \in S$, $x \neq y$, по формуле полной вероятности

$$\begin{gathered}\mathbf{P}\{y \in \left(f_{[k]}\right)^{-r}(x)\} = \sum_{z \in Q_r \setminus \{1\}} \mathbf{P}\{\tau_{f_{[k]}}(y) = z, f^r_{[k]}(y) = x\} + \\ + \sum_{z \in \overline{Q}_r} \mathbf{P}\{\tau_{f_{[k]}}(y) = z, f^r_{[k]}(y) = x\} + \sum_{z=r+1}^{n} \mathbf{P}\{\tau_{f_{[k]}}(y) = z, f^r_{[k]}(y) = x\},\end{gathered} \tag{18}$$

где $\overline{Q}_r = \{1, \ldots, r\} \setminus Q_r$. Заметим, что в случае $z \in Q_r \setminus \{1\}$

$$\mathbf{P}\{\tau_{f_{[k]}}(y) = z, f^r_{[k]}(y) = x, y \in C(G_{f_{[k]}})\} = 0.$$

Тогда для величин, стоящих под первым знаком суммирования в (18), имеем

$$\begin{gathered}\mathbf{P}\{\tau_{f_{[k]}}(y) = z, f^r_{[k]}(y) = x\} = \mathbf{P}\{\tau_{f_{[k]}}(y) = z, f^r_{[k]}(y) = x, y \notin C(G_{f_{[k]}})\} = \\ = \mathbf{P}\{\tau_{f_{[k]}}(y) = z, y \notin C(G_{f_{[k]}})\}\mathbf{P}\{f^r_{[k]}(y) = x \mid \tau_{f_{[k]}}(y) = z, y \notin C(G_{f_{[k]}})\} = \\ = \left(\mathbf{P}\{\tau_{f_{[k]}}(y) = z\} - \mathbf{P}\{y \in C_z(G_{f_{[k]}})\}\right)\frac{(n-2)_{z-2}(z-1)}{(n-1)_{z-1}(z-1)} = \\ = \frac{1}{n-1}\mathbf{P}\{\tau_{f_{[k]}}(y) = z\} - \frac{1}{n-1}\mathbf{P}\{y \in C_z(G_{f_{[k]}})\}.\end{gathered} \tag{19}$$

В случае $z \in \overline{Q}_r \cup \{r+1, \ldots, n\}$ на расположение вершины $y \in \mathcal{K}_{f_{[k]}}(x)$ никаких дополнительных ограничений не накладывается, поэтому

$$\begin{gathered}\mathbf{P}\{\tau_{f_{[k]}}(y) = z, f^r_{[k]}(y) = x\} = \mathbf{P}\{\tau_{f_{[k]}}(y) = z\}\mathbf{P}\{f^r_{[k]}(y) = x \mid \tau_{f_{[k]}}(y) = z\} = \\ = \mathbf{P}\{\tau_{f_{[k]}}(y) = z\}\frac{(n-2)_{z-2}}{(n-1)_{z-1}} = \frac{1}{n-1}\mathbf{P}\{\tau_{f_{[k]}}(y) = z\}.\end{gathered} \tag{20}$$

Подставив выражения (19) и (20) в (18), с учётом равенства [2]

$$\mathbf{P}\{\tau_{f_{[k]}}(y) > z\} = \left(1 - \frac{z}{n}\right)\left(\frac{(n)_z}{n^z}\right)^k$$

и равенства [1]

$$\mathbf{P}\{x \in C_z(G_{f_{[k]}})\} = \frac{1}{n}\left(\frac{(n)_z}{n^z}\right)^k$$

получаем искомую формулу

$$\mathbf{P}\{y \in (f_{[k]})^{-r}(x)\} = \frac{1}{n-1}\left(\sum_{z=2}^{n}\mathbf{P}\{\tau_{f_{[k]}}(y) = z\} - \sum_{z\in Q_r\setminus\{1\}}\mathbf{P}\{y \in C_z(G_{f_{[k]}})\}\right) =$$

$$= \frac{1}{n-1}\mathbf{P}\{\tau_{f_{[k]}}(y) > 1\} - \frac{1}{n-1}\sum_{z\in Q_r\setminus\{1\}}\mathbf{P}\{y \in C_z(G_{f_{[k]}})\} =$$

$$= \frac{1}{n}\left(1 - \frac{1}{n-1}\sum_{z\in Q_r\setminus\{1\}}\left(\frac{(n)_z}{n^z}\right)^k\right).$$

Теорема доказана. ■

**Замечание 4.** Для среднего числа прообразов произвольной фиксированной вершины $x \in S$ в графе $G_{f_{[k]}}$ справедлива цепочка соотношений

$$\mathbf{E}\left|(f_{[k]})^{-r}(x)\right| = \mathbf{E}\sum_{y\in S} I\{y \in (f_{[k]})^{-r}(x)\} = (n-1)\mathbf{P}\{y \in (f_{[k]})^{-r}(x)\} + \mathbf{P}\{x \in (f_{[k]})^{-r}(x)\}.$$

## Заключение

Полученные результаты позволяют описать строение и вероятностные свойства графа $G_{f_{[k]}}$, $k \geqslant 1$, существенно используемые в рамках синтеза и анализа итерационных механизмов защиты информации в части формирования ключей, в принцип функционирования которых заложено применение различных преобразований или источника случайности в каждый отдельный такт работы.

В частности, найдены средние значения мощностей образа исходного множества вершин при действии случайного отображения $f_{[k]}$ и множества вершин, не имеющих прообразов. Получены точные формулы для вероятностей принадлежности указанным множествам фиксированного элемента. Выписаны формулы для распределения длины отрезка апериодичности произвольной фиксированной вершины, не проходящего через заданное множество вершин, вероятностей инцидентности любых двух фиксированных вершин одной компоненте связности и попадания произвольной фиксированной вершины в множество прообразов другой произвольной фиксированной вершины. Вычислена вероятность формирования коллизии произвольной парой фиксированных вершин в случайном графе $G_{f_{[k]}}$.

Автор благодарит А. М. Зубкова за интерес к работе и полезные замечания.

## ЛИТЕРАТУРА

1. *Миронкин В. О.* Распределение длины отрезка апериодичности в графе композиции независимых равновероятных случайных отображений // Математические вопросы криптографии. 2019. Т. 10. № 3. С. 89–99.
2. *Миронкин В. О.* Слои в графе композиции независимых равновероятных случайных отображений // Математические вопросы криптографии. 2020. Т. 11. № 1. С. 101–114.

3. *Зубков А. М., Серов А. А.* Предельная теорема для мощности образа подмножества при композиции случайных отображений // Дискретная математика. 2017. Т. 29. № 1. С. 17–26.
4. *Зубков А. М., Серов А. А.* Оценки среднего размера образа подмножества при композиции случайных отображений // Дискретная математика. 2018. Т. 30. № 2. С. 27–36.
5. *Серов А. А.* Образы конечного множества при итерациях двух случайных зависимых отображений // Дискретная математика. 2015. Т. 27. № 4. С. 133–140.
6. *Dalal A. and Schmutz E.* Compositions of random functions on a finite set // Electr. J. Comb. 2002. V. 9. No. R26. P. 1–7.
7. *Fill J. A.* On compositions of random functions on a finite set // 2002. P. 1–15. `http://www.mts.jhu.edu/~fill/`
8. *Миронкин В. О.* О некоторых вероятностных характеристиках алгоритма выработки ключа “CRYPTOPRO KEY MESHING” // Проблемы информационной безопасности. Компьютерные системы. 2015. № 4. С. 140–146.
9. *Ahmetzyanova L. R., Alekseev E. K., Oshkin I. B., et al.* On the properties of the CTR encryption mode of Magma and Kuznyechik block ciphers with re-keying method based on CryptoPro Key Meshing // Математические вопросы криптографии. 2017. Т. 8. № 2. С. 39–50.
10. *Колчин В. Ф.* Случайные отображения. М.: Наука, 1984. 208 с.
11. *Сачков В. Н.* Вероятностные методы в комбинаторном анализе. М.: Наука, 1978. 288 с.
12. *Harris B.* Probability distributions related to random mapping // Ann. Math. Statist. 1960. V. 31. No. 4. P. 1045–1062.
13. *Flajolet P. and Odlyzko A.* Random mapping statistics // LNCS. 1989. V. 434. P. 329–354.
14. *Колчин В. Ф., Севастьянов Б. А., Чистяков В. П.* Случайные размещения. М.: Наука, 1976. 224 с.

## REFERENCES

1. *Mironkin V. O.* Raspredelenie dlini otrezka aperiodichnosti v grafe kompozicii nezavisimih ravnoveroyatnih sluchaynih otobrazeniy [Distribution of the length of aperiodicity segment in the graph of independent uniform random mappings composition]. Mat. Vopr. Kriptogr., 2019, vol. 10, no. 3, pp. 89–99. (in Russian)
2. *Mironkin V. O.* Sloyi v grafe kompozicii nezavisimih ravnoveroyatnih sluchaynih otobrazeniy [Layers in a graph of the composition of independent uniform random mappings]. Mat. Vopr. Kriptogr., 2020, vol. 11, no. 1, pp. 101–114. (in Russian)
3. *Zubkov A. M. and Serov A. A.* Predelnaya teorema dlya moshnosti obraza podmnozestva pri kompozicii sluchaynih otobrazeniy [Limit theorem for the size of an image of subset under compositions of random mappings]. Discrete Math., 2017, vol. 29, no. 1, pp. 17–26. (in Russian)
4. *Zubkov A. M. and Serov A. A.* Ocenki srednego razmera obraza podmnozestva pri kompozicii sluchaynih otobrazeniy [Estimates of the mean size of the subset image under composition of random mappings]. Discrete Math., 2018, vol. 30, no. 2, pp. 27–36. (in Russian)
5. *Serov A. A.* Obrazi konechnogo mnozestva pri iteraciyah dvuh sluchaynih zavisimih otobrazeniy [Images of a finite set under iterations of two random dependent mappings]. Discrete Math., 2015, vol. 27, no. 4, pp. 133–140. (in Russian)
6. *Dalal A. and Schmutz E.* Compositions of random functions on a finite set. Electr. J. Comb., 2002, vol. 9, no. R26, pp. 1–7.
7. *Fill J. A.* On compositions of random functions on a finite set. 2002, pp. 1–15. `http://www.mts.jhu.edu/~fill/`

8. *Mironkin V. O.* O nekotorih veroyatnostnih harakteristikah algoritma virabotki klucha “CRYPTOPRO KEY MESHING” [On some probabilistic characteristics of key derevation function “CRYPTOPRO KEY MESHING”]. Problemy Informacionnoj Bezopasnosti. Komp’yuternye Sistemy, 2015, no. 4, pp. 140–146. (in Russian)
9. *Ahmetzyanova L. R., Alekseev E. K., Oshkin I. B., et al.* On the properties of the CTR encryption mode of Magma and Kuznyechik block ciphers with re-keying method based on CryptoPro Key Meshing. Mat. Vopr. Kriptogr., 2017, vol. 8, no. 2, pp. 39–50.
10. *Kolchin V. F.* Sluchaynie otobrazeniya [Random Mappings]. Moscow, Nauka Publ., 1984. (in Russian)
11. *Sachkov V. N.* Veroyatnostnie metodi v kombinatornom analize [Probabilistic Methods in Combinatorial Analysis]. Moscow, Nauka Publ., 1978. (in Russian)
12. *Harris B.* Probability distributions related to random mapping. Ann. Math. Statist., 1960, vol. 31, no. 4, pp. 1045–1062.
13. *Flajolet P. and Odlyzko A.* Random mapping statistics. LNCS, 1989, vol. 434, pp. 329–354.
14. *Kolchin V. F., Sevastyanov B. A., and Chistyakov V. P.* Sluchaynie razmesheniya [Random Assignments]. Moscow, Nauka Publ., 1976. (in Russian)

UDC 519.7 DOI 10.17223/20710410/49/2

# METRICAL PROPERTIES OF THE SET OF BENT FUNCTIONS IN VIEW OF DUALITY[1]

A. V. Kutsenko*,**, N. N. Tokareva*

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

**E-mail:** alexandrkutsenko@bk.ru, tokareva@math.nsc.ru

In the paper, we give a review of metrical properties of the entire set of bent functions and its significant subclasses of self-dual and anti-self-dual bent functions. We present results for iterative construction of bent functions in $n+2$ variables based on the concatenation of four bent functions and consider related open problem proposed by one of the authors. Criterion of self-duality of such functions is discussed. It is explored that the pair of sets of bent functions and affine functions as well as a pair of sets of self-dual and anti-self-dual bent functions in $n \geqslant 4$ variables is a pair of mutually maximally distant sets that implies metrical duality. Groups of automorphisms of the sets of bent functions and (anti-)self-dual bent functions are discussed. The solution to the problem of preserving bentness and the Hamming distance between bent function and its dual within automorphisms of the set of all Boolean functions in $n$ variables is considered.

**Keywords:** *Boolean bent function, self-dual bent function, Hamming distance, metrical regularity, automorphism group, iterative construction.*

## 1. Introduction

How much do we know about some cryptographic objects? One way to measure it is to describe what we can do with them. Otherwise, to characterize groups of automorphisms of these objects — separately for each object or together while they form some special class. The question about the group of automorphisms of a set in the Boolean cube necessarily leads us to metrical properties of this set.

That is why we are very interested in *metrical properties* of distinct cryptographic Boolean functions.

The term "bent function" was introduced by Oscar Rothaus in the 1960s [1]. It is known [2], that at the same time Boolean functions with maximal nonlinearity were also studied in the Soviet Union. The term *minimal function*, which is actually a counterpart of a bent function, was proposed by the Soviet scientists Eliseev and Stepchenkov in 1962.

Bent functions have connections with such combinatorial objects as Hadamard matrices and difference sets. Since bent functions have maximum Hamming distance to linear structures and affine functions, they deserve attention for practical applications in symmetric cryptography, in particular for block and stream ciphers. We refer to the survey [3] and monographies of S. Mesnager [4] and N. Tokareva [2] for more information concerning known results and open problems related to bent functions. Results regarding

[1]The work is supported by Mathematical Center in Akademgorodok under agreement No. 075-15-2019-1613 with the Ministry of Science and Higher Education of the Russian Federation and Laboratory of Cryptography JetBrains Research.

the study of metrical properties, in particular, distances between bent functions, one can find in [5].

In this paper we give a review on metrical properties of the entire class of bent function $\mathcal{B}_n$ and its important subclasses — self-dual bent functions $\mathrm{SB}^+(n)$ (i.e. functions such that $f = \widetilde{f}$) and anti-self-dual bent functions $\mathrm{SB}^-(n)$ (i.e. functions such that $f \oplus 1 = \widetilde{f}$), where $\widetilde{f}$ is the dual of $f$. We suppose that the *keys* to the nontrivial and important properties of the class of bent functions are in understanding how does the *duality mapping* $f \to \widetilde{f}$ operate with bent functions. Recall that $\widetilde{\widetilde{f}} = f$ for every bent function $f$. It is important to note that the duality mapping is the *unique* known isometric mapping of the bent functions into themselves that can not be extended to a typical isometry of the whole set of all Boolean functions that preserves bent functions.

On the other hand, the essence of bent functions is expressed in their metrical properties, namely in maximizing distances between them and affine functions. Note that this very idea in more general form is realized in the concept of metrical complement and metrically regular sets. Recall that $\widehat{X}$ is the metrical complement of the set of functions $X$ if it contains all Boolean functions that are on the maximal possible distance from $X$. The set is metrically regular, if $\widehat{\widehat{X}} = X$. There is a some similarity to the self-duality of bent functions, is not it?

Our attention is drawn to automorphism groups of the sets $\mathcal{B}_n$, $\mathcal{A}_n$, $\mathrm{SB}^+(n)$, $\mathrm{SB}^-(n)$ and their metrical properties. Previously, we established that the set of all bent functions $\mathcal{B}_n$ and the set of all affine functions $\mathcal{A}_n$ form a pair of metrically regular sets, i.e. $\widehat{\widehat{\mathcal{B}_n}} = \widehat{\mathcal{A}_n} = \mathcal{B}_n$. Now, we prove the same fact for the classes of self-dual and anti-self-dual functions: they form another such pair of metrically complement functions, i.e. $\widehat{\widehat{\mathrm{SB}^+(n)}} = \widehat{\mathrm{SB}^-(n)} = \mathrm{SB}^+(n)$. In both cases for elements in a pair of metrically regular sets we prove the coincidence of automorphism groups. Thus, $\mathrm{Aut}\,(\mathcal{B}_n) = \mathrm{Aut}\,(\mathcal{A}_n)$ and $\mathrm{Aut}\,\big(\mathrm{SB}^+(n)\big) = \mathrm{Aut}\,\big(\mathrm{SB}^-(n)\big)$. Some other curious properties of bent functions related to their special constructions are discussed.

The paper has the following structure: notation and definitions are in the Section 2. In Section 3, the duality of a bent function is described, including some its important properties and relevant hypothesis proposed by one of the authors (Section 3.1). Some general and metrical properties of the set of bent functions which coincide with their duals, namely self-dual bent functions, are given in Section 3.2. In Section 4, we discuss the iterative construction of bent function in $n+2$ variables based on the concatenation of four bent functions in $n$ variables. The lower bounds on its cardinality and open problem relevant for the set of bent function are in Section 4.1. Criterion of self-duality for bent iterative functions and its corollaries for sign functions together with constructions of self-dual bent functions are discussed in Sections 4.2 and 4.3. In Section 5, the metrical complement of the set of bent functions is studied (Section 5.2) and the results regarding metrical regularity of the set of bent functions and the set of affine functions are given. Metrical complement of the set of (anti-)self-dual bent functions is in Section 5.3. In Section 6, groups of automorphisms of considered sets are studied. The group of automorphisms of the set of bent functions is characterized in Section 6.3 while the (anti-)self-dual case is in Section 6.4. In Section 7, we consider some relations between isometric mappings and the duality of bent function. Isometric mappings which define bijections between the sets of self-dual and anti-self dual bent functions are described in Section 7.1. The Rayleigh quotient of a Boolean function and description of isometric mappings that perserve it or change it for every Boolean function

is given in Section 7.2. The meaning of the Rayleigh quotient in a scope of bent functions is discussed as well.

## 2. Notation

Let $\mathbb{F}_2^n$ be a space of binary vectors of length $n$. A *Boolean function* $f$ in $n$ variables is a map from $\mathbb{F}_2^n$ to $\mathbb{F}_2$. Its *sign function* is $F(x) = (-1)^{f(x)}$, $x \in \mathbb{F}_2^n$. We will also refer to a sign function as to a vector from the set $\{\pm 1\}^{2^n}$:

$$F = (-1)^f = \left((-1)^{f_0}, (-1)^{f_1}, \ldots, (-1)^{f_{2^n-1}}\right) \in \{\pm 1\}^{2^n},$$

where $(f_0, f_1, \ldots, f_{2^n-1}) \in \mathbb{F}_2^{2^n}$ is a truth-table representation of $f$ with arguments given in the lexicographic order. The set of all Boolean functions in $n$ variables is denoted by $\mathcal{F}_n$.

The *algebraic normal form* (ANF, Zhegalkin polynomial) of a Boolean function $f \in \mathcal{F}_n$ is defined as

$$f(x_1, x_2, \ldots, x_n) = \bigoplus_{(i_1, i_2, \ldots, i_n) \in \mathbb{F}_2^n} a_{i_1 i_2 \ldots i_n} x_1^{i_1} x_2^{i_2} \ldots x_n^{i_n},$$

where $a_z \in \mathbb{F}_2$ for any $z \in \mathbb{F}_2^n$ (with the convention $0^0 = 1$). The *algebraic degree* $\deg(f)$ of a Boolean function $f$ is the maximal degree of monomials which occur in its algebraic normal form with nonzero coefficients.

The *Hamming weight* $\mathrm{wt}(x)$ of the vector $x \in \mathbb{F}_2^n$ is the number of nonzero coordinates of $x$. The *Hamming weight* $\mathrm{wt}(f)$ of the function $f \in \mathcal{F}_n$ is the Hamming weight of its vector of values. The *Hamming distance* $\mathrm{dist}(f, g)$ between Boolean functions $f, g$ in $n$ variables is a cardinality of the set $\{x \in \mathbb{F}_2^n : f(x) \oplus g(x) = 1\}$. For $x, y \in \mathbb{F}_2^n$ denote $\langle x, y \rangle = \bigoplus_{i=1}^{n} x_i y_i$. Boolean functions in $n$ variables of the form $f(x) = \langle a, x \rangle \oplus a_0$, $x \in \mathbb{F}_2^n$, where $a_0 \in \mathbb{F}_2$, $a \in \mathbb{F}_2^n$, are called *affine* functions. The set of all affine functions in $n$ variables is denoted by $\mathcal{A}_n$.

The *Walsh – Hadamard transform* (WHT) of a Boolean function $f$ in $n$ variables is an integer valued function $W_f : \mathbb{F}_2^n \to \mathbb{Z}$, defined as

$$W_f(y) = \sum_{x \in \mathbb{F}_2^n} (-1)^{f(x) \oplus \langle x, y \rangle}, \quad y \in \mathbb{F}_2^n.$$

A Boolean function $f$ in an even number $n$ of variables is called *bent* if

$$|W_f(y)| = 2^{n/2}$$

for all $y \in \mathbb{F}_2^n$. The set of all bent functions in $n$ variables is denoted by $\mathcal{B}_n$.

A mapping $\varphi$ of the set of all Boolean functions in $n$ variables to itself is called *isometric* if it preserves the Hamming distance between functions, that is,

$$\mathrm{dist}(\varphi(f), \varphi(g)) = \mathrm{dist}(f, g)$$

for any $f, g \in \mathcal{F}_n$.

Denote, following [6], the orthogonal group of index $n$ over the field $\mathbb{F}_2$ as

$$\mathcal{O}_n = \left\{ L \in \mathrm{GL}(n, \mathbb{F}_2) : L L^{\mathrm{T}} = I_n \right\},$$

where $L^{\mathrm{T}}$ denotes the transpose of $L$ and $I_n$ is the identical matrix of order $n$ over the field $\mathbb{F}_2$.

## 3. The dual of a bent function

From the definition of a bent function it follows that there exists such $\widetilde{f} \in \mathcal{F}_n$ that for any $y \in \mathbb{F}_2^n$ we have

$$W_f(y) = (-1)^{\widetilde{f}(y)} 2^{n/2}.$$

The Boolean function $\widetilde{f}$ defined above is called the *dual* function of the bent function $f$. Thus, for any bent function in $n$ variables its dual Boolean function is uniquely defined. The duality of bent functions was introduced by Dillon [7].

### 3.1. P r o p e r t i e s

Some basic known properties of dual functions are the following [8]:

— Every dual function is a bent function.
— If $\widetilde{f}$ is dual to $f$ and $\widetilde{\widetilde{f}}$ is dual to $\widetilde{f}$, then $\widetilde{\widetilde{f}} = f$.
— The mapping $f \to \widetilde{f}$ which acts on the set of bent functions, preserves the Hamming distance.

There is the following connection between the algebraic degrees of a bent function and its dual [9]:

$$n/2 - \deg(f) \geqslant \frac{n/2 - \deg\left(\widetilde{f}\right)}{\deg\left(\widetilde{f}\right) - 1}.$$

Some results obtained for dual functions can be used in proving the results concerning bent functions, in particular, the connection between ANF coefficients of a bent function and its dual, see [10]:

$$\sum_{x \preccurlyeq y} f(x) = 2^{\mathrm{wt}(y)} - 2^{n/2-1} + 2^{\mathrm{wt}(y)-n/2} \sum_{x \preccurlyeq y \oplus 1} \widetilde{f}(x).$$

One of the most important problem in bent functions is to find the number of them. A new approach to this problem was introduced in [11], see Section 4.1, and the following hypothesis was formulated.

**Hypothesis** (Tokareva, 2011). Any Boolean function in $n$ variables of degree not more than $n/2$ can be represented as the sum of two bent functions in $n$ variables, where $n \geqslant 2$ is an even number.

The review of partial results regarding this problem and also in favour of the Hypothesis one can find in [12]. It was also proved in [13] that

**Theorem 1** [13]. A bent function in $n \geqslant 4$ variables can be represented as the sum of two bent functions in $n$ variables if and only if its dual bent function does.

So, it follows that the mentioned Hypothesis with the decomposition problem, see Section 4.1, can not be considered separately for a bent function and its dual.

It is worth noting that this Hypothesis is a counterpart of the Goldbach's conjecture in number theory unsolved since 1742: any even number $n > 4$ can be represented as the sum of two prime numbers.

Isometric mappings of the set of all Boolean functions in $n$ variables to itself which preserve bentness and the Hamming distance between every bent function and its dual were characterized in [14], namely it was proved that

**Theorem 2** [14]. An isometric mapping $\varphi$ of the set of all Boolean functions in $n$ variables into itself preserves bentness and the Hamming distance between every bent function and its dual if and only if $\varphi$ has form

$$f(x) \longrightarrow f\left(L\left(x \oplus c\right)\right) \oplus \langle c, x\rangle \oplus d, \quad x \in \mathbb{F}_2^n, \tag{1}$$

for some $L \in \mathcal{O}_n$, $c \in \mathbb{F}_2^n$, $\mathrm{wt}(c)$ is even, $d \in \mathbb{F}_2$.

### 3.2. Self-duality

If a bent function $f$ coincides with its dual it is said to be *self-dual*, that is, $f = \widetilde{f}$. A bent function which coincides with the negation of its dual is called an *anti-self-dual*, that is, $f = \widetilde{f} \oplus 1$. The set of (anti-)self-dual bent functions in $n$ variables, according to [15], is denoted by $\mathrm{SB}^+(n)$ $\left(\mathrm{SB}^-(n)\right)$.

Self-dual bent functions were explored in paper of C. Carlet et al. [16] in 2010, where important properties and constructions were given. All equivalence classes of self-dual bent functions in 2, 4 and 6 variables and all quadratic self-dual bent functions in 8 variables with respect to a restricted form of an affine transformation (1), which preserves self-duality, were also presented. Further, equivalence classes of cubic self-dual bent functions in 8 variables with respect to the mentioned above restricted form of affine transformation one can find in [17]. In [15], a classification of quadratic self-dual bent functions was obtained. The upper bound for the cardinality of the set of self-dual bent functions was given in [18]. In [19, 20], one can find new constructions of self-dual bent functions. In papers [21 – 23], several families of self-dual bent functions from involutions were presented. A connection of quaternary self-dual bent functions and self-dual bent Boolean functions was shown in [24]. In [25], it was proved that for $n \geqslant 4$ and any $d \in \{2, 3, \ldots, n/2\}$ there exists a self-dual bent function in $n$ variables of algebraic degree $d$.

In papers [14, 25, 26], metrical properties of the sets of (anti-)self-dual bent functions in $n$ variables were studied. Below we briefly discuss some of them.

Recall that bent functions in $2k$ variables which have a representation

$$f(x, y) = \langle x, \pi(y)\rangle \oplus g(y), \quad x, y \in \mathbb{F}_2^k,$$

where $\pi : \mathbb{F}_2^k \to \mathbb{F}_2^k$ is a permutation and $g$ is a Boolean function in $k$ variables, form the well known *Maiorana — McFarland class* of bent functions [27]. Necessary and sufficient conditions of (anti-)self-duality of bent functions from Maiorana — McFarland class are known from [16]. Let the denotion $\mathrm{SB}^+_{\mathcal{M}}(n)$ stands for the set of self-dual Maiorana — McFarland bent functions and $\mathrm{SB}^-_{\mathcal{M}}(n)$ for the set of anti-self-dual ones both in $n$ variables. In [26], the set of possible Hamming distance between such self-dual bent functions was found.

**Theorem 3** [26]. Let $n \geqslant 4$ and $f, g \in \mathrm{SB}^+_{\mathcal{M}}(n) \cup \mathrm{SB}^-_{\mathcal{M}}(n)$, then

$$\mathrm{dist}(f, g) \in \left\{2^{n-1}, 2^{n-1}\left(1 \pm \frac{1}{2^r}\right), r = 0, 1, ..., n/2 - 1\right\}.$$

Moreover, if either $f, g \in \mathrm{SB}^+_{\mathcal{M}}(n)$ or $f, g \in \mathrm{SB}^-_{\mathcal{M}}(n)$, then all distances are attainable, and for any pair $f \in \mathrm{SB}^+_{\mathcal{M}}(n)$ and $g \in \mathrm{SB}^-_{\mathcal{M}}(n)$ it holds $\mathrm{dist}(f, g) = 2^{n-1}$.

By analysis of the set of distances from Theorem 3, the minimal Hamming distance between considered functions can be obtained.

**Corollary 1.** Let $n \geqslant 4$, then the minimal Hamming distance between (anti-)self-dual Maiorana — McFarland bent functions is equal to $2^{n-2}$.

Moreover, since the minimal Hamming distance between quadratic Boolean functions in $n$ variables (which correspond to codewords of the $\mathrm{RM}(2,n)$ code) is at least $2^{n-2}$ [28], the following fact holds.

**Corollary 2.** Let $n \geqslant 4$, then the minimal Hamming distance between quadratic bent functions can be attained on (anti-)self-dual Maiorana — McFarland bent functions.

It is known that the minimal Hamming distance between bent functions in $n$ variables is $2^{n/2}$ [5]. In [25], it was proved that this extremal value can be attained on (anti-)self-dual bent functions.

**Theorem 4** [25]. Let $n \geqslant 4$, then the minimal Hamming distance between distinct (anti-)self-dual bent functions in $n$ variables is equal to $2^{n/2}$.

In the case $n = 2$, there are only two self-dual Maiorana — McFarland bent functions, namely $f_1(x_1, x_2) = x_1x_2$ and $f_2(x_1, x_2) = x_1x_2 \oplus 1$, and two anti-self-dual Maiorana — McFarland bent functions, namely $g_1(x_1, x_2) = x_1x_2 \oplus x_1 \oplus x_2$ and $g_2(x_1, x_2) = x_1x_2 \oplus x_1 \oplus \oplus x_2 \oplus 1$. It is clear that $\mathrm{dist}(f_1, g_1) = \mathrm{dist}(f_2, g_2) = 4 = 2^n$ and $\mathrm{dist}(f_1, g_2) = \mathrm{dist}(f_2, g_1) = = 2 = 2^{n-1}$.

## 4. Iterative construction $\mathcal{BI}$

Let $f_0, f_1, f_2, f_3$ be Boolean functions in $n$ variables. Consider a Boolean function $g$ in $n + 2$ variables which is defined as

$$g(00, x) = f_0(x), \quad g(01, x) = f_1(x), \quad g(10, x) = f_2(x), \quad g(11, x) = f_3(x), \quad x \in \mathbb{F}_2^n.$$

It is known (Preneel et al., 1991; see also [11, 29]) that under condition $f_0, f_1, f_2, f_3 \in \mathcal{B}_n$ the mentioned function $g$ is a bent function in $n + 2$ variables if and only if

$$\widetilde{f}_0 \oplus \widetilde{f}_1 \oplus \widetilde{f}_2 \oplus \widetilde{f}_3 = 1,$$

that gives the construction of a bent function in $n + 2$ variables through the concatenation of vectors of values of four bent functions in $n$ variables [30].

Following N. Tokareva [11], we will refer to bent functions obtained by this construction as *bent iterative functions* ($\mathcal{BI}$) and denote the set of such bent functions in $n$ variables by $\mathcal{BI}_n$.

In [31], the comparison of cardinalities of different known iterative constructions of bent functions in $n \leqslant 10$ variables was presented and the class $\mathcal{BI}$ had the biggest cardinality among them.

According to [29], there exist bent functions from Maiorana — McFarland class [27] and from the class $\mathcal{PS}$ (Partial Spreads) [7] that can not be represented as bent iterative functions. Also, from paper [32] on nonnormal bent functions, it follows that there exist bent functions in $\mathcal{BI}_n$ that are nonequivalent to Maiorana — McFarland bent functions.

### 4.1. Lower bounds on the cardinality and related open problem

In paper [11], some possible methods for calculating the number of bent iterative functions were shown.

**Theorem 5** [11]. For any even $n \geqslant 4$

$$|\mathcal{BI}_n| = \sum_{f' \in \mathcal{B}_{n-2}} \sum_{f'' \in \mathcal{B}_{n-2}} |(\mathcal{B}_{n-2} \oplus f') \cap (\mathcal{B}_{n-2} \oplus f'')| .$$

Denote $X_n = \{f \oplus h : f, h \in \mathcal{B}_n\}$ and consider the system $\{C_f : f \in \mathcal{B}_n\}$ of its subsets defined as $C_f = \mathcal{B}_n \oplus f$. So

$$X_n = \bigcup_{f \in \mathcal{B}_n} C_f.$$

Let $\psi$ be an element of $X_n$. The number of subsets $C_f$ that cover $\psi$, according to [11], is called *multiplicity* of $\psi$ and is denoted by $m(\psi)$. One can notice that if $\psi$ is covered by $C_f$, then it is covered by any set $C_{f'}$, where $f'$ is obtained from $f$ by adding an affine function.

In [11], the exact number of bent iterative functions through the multiplicities was obtained.

**Theorem 6** [11]. For any even $n \geqslant 2$,

$$|\mathcal{BI}_{n+2}| = \sum_{\psi \in C_f} m^2(\psi).$$

So in order to evaluate $|\mathcal{BI}_{n+2}|$ (and then $|\mathcal{B}_{n+2}|$) we have to study the set $X_n$ and the distribution of multiplicities for its elements. Such an analysis, as shown in [11], gives the following lower bound.

**Theorem 7** [11]. For any even $n \geqslant 2$,

$$\frac{|\mathcal{B}_{n+2}|^4}{|X_n|} \leqslant |\mathcal{BI}_{n+2}| \leqslant |\mathcal{B}_{n+2}|.$$

Thus, for calculating the exact number of bent iterative functions, one has to study the structure of the set $X_n$. So we come to a new problem statement.

**Open problem: bent sum decomposition** (Tokareva, 2011). What Boolean functions can be represented as the sum of two bent functions in $n$ variables? How many such representations does a Boolean function admit?

The related Hypothesis was previuosly mentioned in the Section 3.1.

### 4.2. Self-dual bent iterative functions

The set of (anti-)self-dual bent functions from $\mathcal{BI}_n$ is further denoted by $\mathrm{SB}^{+}_{\mathcal{BI}}(n)$ $\left(\mathrm{SB}^{-}_{\mathcal{BI}}(n)\right)$.

In paper [25], the necessary and sufficient conditions of self-duality of bent iterative functions were studied, namely, the following result was obtained: taking constant function $h$, we can obtain two constructions of self-dual bent iterative functions in $n+2$ variables.

**Theorem 8** [25]. Let $g \in \mathcal{BI}_{n+2}$. Then $g$ is self-dual bent if and only if there exists such pair of functions $g_1, g_2 \in \mathcal{B}_n$, that

$$\begin{aligned}
f_0 &= (g_1 \oplus g_2)\, h \oplus g_1 = \widetilde{g}_2,\\
f_1 &= (g_1 \oplus g_2)\, h \oplus g_2 = \widetilde{g_1 \oplus h},\\
f_2 &= (g_1 \oplus g_2)\, h \oplus g_2 \oplus h = \widetilde{g}_1,\\
f_3 &= (g_1 \oplus g_2)\, h \oplus g_1 \oplus h \oplus 1 = \widetilde{g_2 \oplus h} \oplus 1,
\end{aligned}$$

where the function $h \in \mathcal{F}_n$ is uniquely defined by a pair of bent functions $g_1, g_2$, namely:

$$h = g_1 \oplus \widetilde{g}_1 \oplus g_2 \oplus \widetilde{g}_2.$$

Two iterative constructions of self-dual bent functions immeditely follow from Theorem 8, as it was shown in [25].

**Corollary 3.** Functions

$$f'(y_1, y_2, x) = (y_1 \oplus y_2)\left(f(x) \oplus \widetilde{f}(x)\right) \oplus f(x) \oplus y_1 y_2,$$
$$f''(y_1, y_2, x) = (y_1 \oplus y_2)(\varphi(x) \oplus \omega(x)) \oplus \varphi(x) \oplus \alpha_1 y_1 \oplus \alpha_2 y_2 \oplus y_1 y_2,$$

where $y_1, y_2, \alpha_1, \alpha_2 \in \mathbb{F}_2$, $\alpha_1 \oplus \alpha_2 = 1$, $x \in \mathbb{F}_2^n$, $f \in \mathcal{B}_n$, $\varphi \in \mathrm{SB}^+(n)$, $\omega \in \mathrm{SB}^-(n)$, are self-dual bent functions in $n+2$ variables.

The first construction (for $f'$) was earlier presented in [16] as an example of the construction which uses the indirect sum of bent functions, see [8]. It is worth noting that the second construction (for $f''$) can also be obtained from indirect sum of bent functions.

Since these constructions do not intersect, the sum of their cardinalities provides a lower bound for the cardinality of the set of self-dual bent iterative functions [25].

**Corollary 4.** $|\mathcal{B}_{n-2}| + \left|\mathrm{SB}^+(n-2)\right|^2 \leqslant \left|\mathrm{SB}^+_{\mathcal{BI}}(n)\right| \leqslant |\mathcal{B}_{n-2}|^2$.

### 4.3. The dimension of linear span of sign functions of self-dual bent functions

Let $H_n = H_1^{\otimes n}$ be the $n$-fold tensor product of the matrix $H_1$ with itself, where

$$H_1 = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}.$$

It is known the Hadamard property of this matrix:

$$H_n H_n^{\mathrm{T}} = 2^n I_{2^n}.$$

Denote $\mathcal{H}_n = 2^{-n/2} H_n$. In terms of sign functions, the function $f \in \mathcal{F}_n$ is bent if for its sign function $F$ it holds $\mathcal{H}_n F \in \{\pm 1\}^{2^n}$.

Recall that a non-zero vector $v \in \mathbb{C}^n$ is called an *eigenvector* of a square $n \times n$ matrix $A$ attached to the eigenvalue $\lambda \in \mathbb{C}$ if $Av = \lambda v$. A linear span of eigenvectors attached to the eigenvalue $\lambda$ is called an *eigenspace* associated with $\lambda$. Consider a linear mapping $\psi : \mathbb{C}^n \to \mathbb{C}^n$ represented by a $n \times n$ complex matrix $A$. A *kernel* of $\psi$ is the set

$$\mathrm{Ker}\,(\psi) = \{x \in \mathbb{C}^n : Ax = \mathbf{0} \in \mathbb{C}^n\},$$

where $\mathbf{0}$ is a zero element of the space $\mathbb{C}^n$.

From the definition of self-duality it follows that sign function of any self-dual bent function is the eigenvector of $\mathcal{H}_n$ attached to the eigenvalue 1, that is an element from the subspace $\mathrm{Ker}\,(\mathcal{H}_n - I_{2^n}) = \mathrm{Ker}\left(H_n - 2^{n/2} I_{2^n}\right)$. The same holds for a sign function of any anti-self-dual bent function, which obviously is an eigenvector of $\mathcal{H}_n$ attached to the eigenvalue $(-1)$, that is, an element from the subspace $\mathrm{Ker}\,(\mathcal{H}_n + I_{2^n}) = \mathrm{Ker}\left(H_n + 2^{n/2} I_{2^n}\right)$.

In [16], an orthogonal decomposition of $\mathbb{R}^{2^n}$ in eigenspaces of $H_n$ was given:

$$\mathbb{R}^{2^n} = \mathrm{Ker}\left(H_n + 2^{n/2} I_{2^n}\right) \oplus \mathrm{Ker}\left(H_n - 2^{n/2} I_{2^n}\right), \tag{2}$$

where the symbol $\oplus$ denotes a direct sum of subspaces.

It is known that

$$\dim\left(\mathrm{Ker}\left(H_n + 2^{n/2} I_{2^n}\right)\right) = \dim\left(\mathrm{Ker}\left(H_n - 2^{n/2} I_{2^n}\right)\right) = 2^{n-1},$$

where $\dim(V)$ is the dimension of the subspace $V \subseteq \mathbb{R}^{2^n}$. Moreover, from symmetricity of $\mathcal{H}_n$ it follows that the subspaces $\mathrm{Ker}\left(H_n - 2^{n/2} I_{2^n}\right)$ and $\mathrm{Ker}\left(H_n + 2^{n/2} I_{2^n}\right)$ are mutually orthogonal.

In [25], it was proved that

**Theorem 9** [25]. If $n \geqslant 4$, then:

— among sign functions of self-dual bent functions in $n$ variables there exists a basis of the eigenspace of the matrix $H_n$ attached to the eigenvalues 1, that is, the subspace $\mathrm{Ker}\left(H_n - 2^{n/2} I_{2^n}\right)$;

— among sign functions of anti-self-dual bent functions in $n$ variables there exists a basis of the eigenspace of the matrix $H_n$ attached to the eigenvalues $(-1)$, that is, the subspace $\mathrm{Ker}\left(H_n + 2^{n/2} I_{2^n}\right)$.

It is worth notice that there exists an example of basis which consists of sign functions of self-dual bent iterative functions provided by two constructions of self-dual bent iterative functions obtained by Theorem 8. Given the basis for self-dual case, the basis for anti-self-dual case can be obtained by using one of bijections from Theorem 20.

## 5. Metrical complement and regularity

In this section, we give results regarding notable metrical property of a subset of Boolean cube called metrical regularity. The sets of affine Boolean functions and bent functions possess it. The sets of self-dual and anti-self-dual bent functions in $n \geqslant 4$ variables are also mutually maximally distant. That implies metrical *duality*, in some sence, between the considered pairs of subsets of Boolean functions.

Regarding that, some essential and intriguing questions arise: for instance, are there any pairs of metrically regular subsets inside the metrically regular set of bent functions in $n$ variables? If additionally, in order to exclude some trivial cases, we consider only the subsets which include functions together with their negations, the maximal Hamming distance from the considered sets is at most $2^{n-1}$. Are there any pairs of metrically regular subsets with additional mentioned requirement such that the distance between them is exactly $2^{n-1}$, that is, they would be extreme?

### 5.1. Definitions

Let $X \subseteq \mathbb{F}_2^n$ be an arbitrary set and let $y \in \mathbb{F}_2^n$ be an arbitrary vector. Define the *distance* between $y$ and $X$ as $\mathrm{dist}(y, X) = \min\limits_{x \in X} \mathrm{dist}(y, x)$. The *maximal distance* from the set $X$ is

$$\mathrm{d}(X) = \max_{y \in \mathbb{F}_2^n} \mathrm{dist}(y, X).$$

In coding theory this number is also known as the *covering radius* of the set $X$. A vector $z \in \mathbb{F}_2^n$ is called *maximally distant* from a set $X$ if $\mathrm{dist}(z, X) = \mathrm{d}(X)$. The set of all maximally distant vectors from the set $X$ is called the *metrical complement* of the set $X$ and is denoted by $\widehat{X}$ [33]. A set $X$ is said to be *metrically regular* if $\widehat{\widehat{X}} = X$. Define, following N. Tokareva [2], a subset of Boolean functions to be *metrically regular* if the set of corresponding vectors of values is metrically regular.

Sets of functions which have maximum distance from partition set functions were studied in [34], it was shown that partition set functions defined by some partition are mutually maximally distant sets. Lower bound on size of the largest metrically regular subset of the Boolean cube was studied in [35].

### 5.2. The set of bent functions

Let $\mathrm{GA}(n)$ denote an affine group.

**Proposition 1.** Any isometric mapping of the form

$$f(x) \longrightarrow f\left(Ax \oplus b\right) \oplus \langle c, x\rangle \oplus d,$$

where $A \in \mathrm{GL}(n)$, $b, c \in \mathbb{F}_2^n$, $d \in \mathbb{F}_2$, preserves bentness.

In [36], the following theorem was proved.

**Theorem 10** [36]. For each non-affine Boolean function $h \in \mathcal{F}_n$, there exists a bent function $f \in \mathcal{B}_n$ such that $f \oplus h$ is not bent.

From Proposition 1 and Theorem 10 it follows that the set of bent functions is closed under addition of affine Boolean functions only. This fact implies that the affine functions are precisely all Boolean functions which are at the maximum distance from the class of bent functions. Namely, in [36] it was shown that

**Theorem 11** [36]. A Boolean function in $n$ variables is

— a bent function if and only if it has the maximal possible distance $2^{n-1} - 2^{n/2-1}$ to the set of all affine functions, that is it is an element of $\widehat{\mathcal{A}_n}$;

— an affine function if and only if it has the maximal possible distance $2^{n-1} - 2^{n/2-1}$ to the set of all bent functions, that is it is an element of $\widehat{\mathcal{B}_n}$.

Thus, from the results given in [36], it follows that there exists a *duality*, in some sense, between the definitions of bent functions and affine functions. In particular, we obtain metrical regularity of the sets of affine functions and bent functions.

**Corollary 5.**

1) The set $\mathcal{A}_n$ of all affine Boolean functions in $n$ variables is metrically regular.
2) The set $\mathcal{B}_n$ of all bent functions in $n$ variables is metrically regular.

### 5.3. The set of (anti-)self-dual bent functions

For any (anti-)self-dual bent function $f \in \mathrm{SB}^+(n)$ its negation $f \oplus 1$ is also (anti-)self-dual bent [16, 17]. Moreover, from the results presented in [14], it follows the counterpart of Theorem 10 for the (anti-)self-dual case, namely:

**Theorem 12.** For each non-constant Boolean function $h \in \mathcal{F}_n$ there exists a self-dual bent function $f \in \mathrm{SB}^+(n)$ such that $f \oplus h$ is not self-dual bent. Anti-self-dual bent functions possess the same property.

Thus, it follows that the set of (anti-)self-dual bent functions is closed only under addition of 1, that is, taking the negation of the function.

From the fact that considered set is closed under addition of 1, it follows that the maximal Hamming distance from the set $\mathrm{SB}^+(n)$ is at most $2^{n-1}$. It was proved by Carlet et al. in [16] that the Hamming distance between any pair of self-dual and anti-self-dual bent functions, both in $n$ variables, is equal to $2^{n-1}$. So we have

$$\mathrm{d}\left(\mathrm{SB}^+(n)\right) = 2^{n-1},$$

and all anti-self-dual bent functions in $n$ variables belong to the metrical complement of the set of self-dual bent functions in $n$ variables.

In paper [25], the metrical complement of the set of (anti-)self-dual bent functions in $n \geqslant 4$ variables was completely characterized by using the orthogonal decomposition (2) and existence of the basis provided by the Theorem 9.

**Theorem 13** [25]. Let $n \geqslant 4$, then a Boolean function in $n$ variables is:

— self-dual bent if and only if it has the maximal possible distance $2^{n-1}$ to the set of all anti-self-dual bent functions, that is, it is an element of $\widehat{\mathrm{SB}^-(n)}$;

— anti-self-dual bent if and only if it has the maximal possible distance $2^{n-1}$ to the set of all self-dual bent functions, that is, it is an element of $\widehat{\mathrm{SB}^+(n)}$.

As for the pair of the sets of bent functions and affine functions, it follows that there also exists a *duality* between the sets of self-dual and anti-self-dual bent functions in $n \geqslant 4$ variables.

The case $n = 2$ was considered explicitly and it appeared that both $\mathrm{SB}^+(2)$ and $\mathrm{SB}^-(2)$ are metrically regular sets. From that and the Theorem 13 it follows

**Corollary 6.**

1) The set $\mathrm{SB}^+(n)$ of all self-dual bent functions in $n$ variables is metrically regular.
2) The set $\mathrm{SB}^-(n)$ of all anti-self-dual bent functions in $n$ variables is metrically regular.

## 6. The group of automorphisms

Study of automorphism groups of mathematical objects deserves attention since these groups are closely connected with the structure of the objects. There exists a natural question: how groups of automorphisms of two mathematical objects, one of which is embedded to another one, are related.

An example of such a problem statement is the set of bent functions in $n$ variables and one of its significant subclasses which consisits of self-dual bent functions in $n$ variables.

It is also worth mentioning that the complexity of classification of combinatorial objects depends on generality of the approach. Consequently, the question "*if the common approach to classify* (*self-dual*) *bent functions is the most general within automorphisms of the set of Boolean functions*", arises naturally.

### 6.1. Isometric mappings and automorphism groups

Recall that a mapping $\varphi$ of the set of all Boolean functions in $n$ variables to itself is called *isometric* if it preserves the Hamming distance between functions. Following [14], denote the set of all isometric mappings of the set of all Boolean functions in $n$ variables to itself by $\mathcal{I}_n$.

It is known (A. A. Markov, 1956) that every isometric mapping of all Boolean functions in $n$ variables to itself has the unique representation of the form

$$f(x) \longrightarrow f(\pi(x)) \oplus g(x), \tag{3}$$

where $\pi$ is a permutation on the set $\mathbb{F}_2^n$ and $g \in \mathcal{F}_n$ [37]. The mapping of this form is denoted by $\varphi_{\pi,g} \in \mathcal{I}_n$.

The *group of automorphisms* of a fixed subset $M \subseteq \mathcal{F}_n$ is the group of isometric mappings of the set of all Boolean functions in $n$ variables to itself preserving the set $M$. It is denoted by $\mathrm{Aut}\,(M)$.

### 6.2. Matrix representation

For the number $k \in \{0, 1, \ldots, 2^n - 1\}$, denote by $\mathbf{v}_k \in \mathbb{F}_2^n$ its binary representation.

Recall that a square matrix is called *monomial* (or *generalized permutation matrix*) if it has exactly one nonzero entry in each row and each column.

The following one-to-one correspondence between the set $\mathcal{I}_n$ and the set of monomial matrices of order $2^n$ with nonzero elements from the set $\{\pm 1\}$ was used in [14]. In more

detail, let $\varphi_{\pi,g} \in \mathcal{I}_n$ be an arbitrary isometric mapping. Then, for any $f \in \mathcal{F}_n$ and its sign function

$$F = \left((-1)^{f(\mathbf{v}_0)}, (-1)^{f(\mathbf{v}_1)}, \ldots, (-1)^{f(\mathbf{v}_{2^n-1})}\right) \in \{\pm 1\}^{2^n},$$

the sign function

$$F' = \left((-1)^{f'(\mathbf{v}_0)}, (-1)^{f'(\mathbf{v}_1)}, \ldots, (-1)^{f'(\mathbf{v}_{2^n-1})}\right) \in \{\pm 1\}^{2^n}$$

of $f' = \varphi_{\pi,g}(f) \in \mathcal{F}_n$ can be expressed as $F' = AF$, where $A$ is the $2^n \times 2^n$ monomial matrix, constructed by the permutation $\pi$ and the function $g$:

$$i \begin{pmatrix} & & & & j & & & \\ & & & & \vdots & & & \\ & & & & 0 & & & \\ & & & & \vdots & & & \\ \ldots & 0 & \ldots & & (-1)^{g(\mathbf{v}_{i-1})} & \ldots & 0 & \ldots \\ & & & & \vdots & & & \\ & & & & 0 & & & \\ & & & & \vdots & & & \end{pmatrix},$$

in which in the $i$-th row a nonzero element $(-1)^{g(\mathbf{v}_{i-1})}$ is in the $j$-th column, where $(j-1)$ is a number with binary representation $\pi(\mathbf{v}_{i-1})$. So the $i$-th component of $F' = AF$ is equal to

$$(-1)^{f'(\mathbf{v}_{i-1})} = (-1)^{f(\pi(\mathbf{v}_{i-1}))}(-1)^{g(\mathbf{v}_{i-1})} = (-1)^{f(\pi(\mathbf{v}_{i-1}))\oplus g(\mathbf{v}_{i-1})}$$

for any $i \in \{1, 2, \ldots, 2^n\}$, that is, equivalent to $f'(x) = f(\pi(x)) \oplus g(x)$, $x \in \mathbb{F}_2^n$.

6.3. The group of automorphisms of the set of bent functions

Some attempts to determine the automorphism group of a given bent function were undertaken by U. Dempwolff in 2006 [38]. Results were presented in terms of elementary Abelian Hadamard difference sets (equivalently, bent functions).

A natural question whether there exist isometric mappings of Boolean functions into itself, distinct from those mentioned in Proposition 1, which preserve the class of bent function, was completely solved in paper [39]. It was proved that there were no other mappings possessing such a property. Namely, by using the Theorem 11 in view of the duality, the following coincidence was shown.

**Theorem 14** [39]. $\mathrm{Aut}(\mathcal{B}_n) = \mathrm{Aut}(\mathcal{A}_n)$.

The group of automorphisms of the set of all affine functions in $n$ variables consists, as it is well known, of mappings of the form (3) with affine permutation $\pi$ and affine shift $g$, see, for example, [28]. Note that the set of all affine functions in $n$ variables forms a group isomorphic to $\mathbb{F}_2^{n+1}$. Let the symbol $\ltimes$ stands for the semidirect product, then the result is formulated as follows.

**Theorem 15** [39]. $\mathrm{Aut}(\mathcal{B}_n) = \mathrm{GA}(n) \ltimes \mathbb{F}_2^{n+1}$.

These results imply the non-existence of a more general approach to equivalence of bent functions than that on the base of isometric mappings.

### 6.4. The group of automorphisms of the set of (anti-)self-dual bent functions

In [16], the following problem was pointed.

**Open question** (Carlet, Danielson, Parker, Solé, 2010): to find mappings preserving self-duality, distinct from the known ones, or give a proof that there are no more.

In [14], this question was resolved within isometric mappings of the set of all Boolean functions in $n \geqslant 4$ variables into itself.

First, there is the problem of how the sets of isometric mapping preserving self-duality and anti-self-duality or, in other words, groups of automorphisms of the sets $\mathrm{SB}^+(n)$ and $\mathrm{SB}^-(n)$ are related. This problem was solved in [14], where with a use of the orthogonal decomposition (2) and the basis from the Theorem 9 it was proved

**Theorem 16** [14]. If $n \geqslant 4$, then $\mathrm{Aut}\left(\mathrm{SB}^+(n)\right) = \mathrm{Aut}\left(\mathrm{SB}^-(n)\right)$.

In [14], the criterion of preserving self-duality was also presented.

**Theorem 17** [14]. If $n \geqslant 4$, then isometric mapping $\varphi_{\pi,g}$ belongs to $\mathrm{Aut}\left(\mathrm{SB}^+(n)\right)$ if and only if, for any $x, y \in \mathbb{F}_2^n$, it holds

$$\langle \pi(x), y\rangle \oplus g(x) = \left\langle x, \pi^{-1}(y)\right\rangle \oplus g\left(\pi^{-1}(y)\right).$$

In matrix terms the criterion can be formulated as $A\mathcal{H}_n = \mathcal{H}_n A$, where $A$ is the matrix which represents the mapping $\varphi_{\pi,g}$.

The problem of characterization mappings which preserve self-duality was studied in [16, 17], where it was shown that the mapping (1) preserves self-duality of a bent function, in other words, it is an element of $\mathrm{Aut}\left(\mathrm{SB}^+(n)\right)$. It is obvious that this mapping is isometric and corresponds to $\varphi_{\pi,g} \in \mathcal{I}_n$ with

$$\pi(x) = L\left(x \oplus c\right),\ g(x) = \langle c, x\rangle \oplus d,\ x \in \mathbb{F}_2^n,$$

where $L \in \mathcal{O}_n$, $c \in \mathbb{F}_2^n$, $\mathrm{wt}(c)$ is even, $d \in \mathbb{F}_2$. The group which consists of mappings of such form is called an *extended orthogonal group* and denoted by $\overline{\mathcal{O}}_n$ [17, 40]. It is known that this group is a subgroup of $\mathrm{GL}\left(n+2, \mathbb{F}_2\right)$ [17].

In paper [14], known results were generalized within isometric mappings from the set $\mathcal{I}_n$ for $n \geqslant 4$. Namely, by using the criterion from Theorem 17 and the matrix representation of isometric mappings (see Section 6.2), it was proved that the desired group of automorphisms coincides with the extended orthogonal group.

**Theorem 18** [14]. For $n \geqslant 4$,

$$\mathrm{Aut}\left(\mathrm{SB}^+(n)\right) = \mathrm{Aut}\left(\mathrm{SB}^-(n)\right) = \overline{\mathcal{O}}_n.$$

It follows that the classification of self-dual bent functions in $n \geqslant 4$ variables based on the restricted form of affine equivalence proposed in [16, 17] is the most general isometric mapping of the set of all Boolean functions in $n$ variables into itself.

## 7. Isometric mappings and duality

In this Section, we discuss results from [14] on characterization of isometric mappings which define bijections between self-dual and anti-self dual bent functions, and description of isometric mappings which preserve or change the sign of the Rayleigh quotient of a Boolean function.

### 7.1. Isometric bijections between self-dual and anti-self-dual bent functions

It is known [16] that there exists a bijection between $\mathrm{SB}^+(n)$ and $\mathrm{SB}^-(n)$, based on the decomposition of sign functions of (anti-)self-dual bent functions. Also, note that from the existence of such bijection it follows that $\left|\mathrm{SB}^+(n)\right| = \left|\mathrm{SB}^-(n)\right|$.

Namely, let $(Y, Z) \in \{\pm 1\}^{2^n}$, where $Y, Z \in \{\pm 1\}^{2^{n-1}}$, be a sign function for some $f \in \mathrm{SB}^+(n)$. Then a vector $(Z, -Y) \in \{\pm 1\}^{2^n}$ is a sign function for some function from $\mathrm{SB}^-(n)$. In terms of isometric mappings, this transformation can be represented as

$$f(x) \longrightarrow f\left(x \oplus c\right) \oplus \langle c, x \rangle,$$

where $c = (1, 0, 0, \ldots, 0) \in \mathbb{F}_2^n$.

In [15], it was mentioned that the more general form of this mapping

$$f(x) \longrightarrow f\left(x \oplus c\right) \oplus \langle c, x \rangle,$$

where $c \in \mathbb{F}_2^n$, $\mathrm{wt}(c)$ is odd, is a bijection between $\mathrm{SB}^+(n)$ and $\mathrm{SB}^-(n)$. It is obvious that this mapping is an element from $\mathcal{I}_n$.

In [14], these results were generalized within isometric mappings from the set $\mathcal{I}_n$ for $n \geqslant 4$.

The criterion of bijectivity between self-dual and anti-self-dual bent functions was obtained in [14] with a use of the orthogonal decomposition (2) and the basis from the Theorem 9.

**Theorem 19** [14]. Let $n \geqslant 4$, then isometric mapping $\varphi_{\pi,g} \in \mathcal{I}_n$ is a bijection between $\mathrm{SB}^+(n)$ and $\mathrm{SB}^-(n)$ if and only if, for any $x, y \in \mathbb{F}_2^n$, it holds

$$\langle \pi(x), y \rangle \oplus g(x) = \left\langle x, \pi^{-1}(y) \right\rangle \oplus g\left(\pi^{-1}(y)\right) \oplus 1.$$

By using this criterion, in [14] the general form of considered isometric bijections was found.

**Theorem 20** [14]. For $n \geqslant 4$, isometric mapping $\varphi_{\pi,g} \in \mathcal{I}_n$ is a bijection between $\mathrm{SB}^+(n)$ and $\mathrm{SB}^-(n)$ if and only if

$$\pi(x) = L\left(x \oplus c\right), \ g(x) = \langle c, x \rangle \oplus d, \ x \in \mathbb{F}_2^n,$$

where $L \in \mathcal{O}_n$, $c \in \mathbb{F}_2^n$, $\mathrm{wt}(c)$ is odd, $d \in \mathbb{F}_2$.

Thus, from Theorems 18 and 20 we can conclude that if we take a mapping from the group $\overline{\mathcal{O}}_n$ and replace the vector $c \in \mathbb{F}_2^n$ by a binary vector of length $n$ with an odd Hamming weight, then we switch the mapping from the "automorphism mode" to the "bijection mode" between the sets $\mathrm{SB}^+(n)$ and $\mathrm{SB}^-(n)$.

### 7.2. Isometric mappings and the Rayleigh quotient

In [16], the *Rayleigh quotient* $S_f$ of a Boolean function $f \in \mathcal{F}_n$ was defined as

$$S_f = \sum_{x,y \in \mathbb{F}_2^n} (-1)^{f(x) \oplus f(y) \oplus \langle x, y \rangle} = \sum_{y \in \mathbb{F}_2^n} (-1)^{f(y)} W_f(y).$$

In a scope of bent functions, the Rayleigh quotient characterizes the Hamming distance between a bent function and its dual. Indeed, let $f \in \mathcal{B}_n$, then

$$\mathrm{dist}(f, \widetilde{f}) = 2^{n-1} - \frac{1}{2^{n/2+1}} S_f = 2^{n-1} - \frac{1}{2} N_f.$$

In [16], it was proved that, for any $f \in \mathcal{F}_n$, the absolute value of $S_f$ is at most $2^{3n/2}$ with equality if and only if $f$ is self-dual $\left(+2^{3n/2}\right)$ and anti-self-dual $\left(-2^{3n/2}\right)$ bent function. That is, the maximum (minimum) value of the Rayleigh quotient of a Boolean function in an even number of variables is attainable on self-dual (anti-self-dual) bent functions and only them, thus providing a criterion for (anti-)self-duality in terms of the Rayleigh quotient values.

In [40], the operations on Boolean functions that preserve bentness and the Rayleigh quotient were given. Namely, it was proved that, for any $f \in \mathcal{B}_n$, $L \in \mathcal{O}_n$, $c \in \mathbb{F}_2^n$, $d \in \mathbb{F}_2$, the functions $g, h \in \mathcal{B}_n$ defined as $g(x) = f(Lx) \oplus d$ and $h(x) = f(x \oplus c) \oplus \langle c, x \rangle$ provide $N_g = N_f$ and $N_h = (-1)^{\langle c,c \rangle} N_f$.

The mentioned operations are isometric mappings from $\mathcal{I}_n$. The complete characterization of isometric mappings that preserve the Rayleigh quotient as well as change it was given in [14].

**Theorem 21** [14]. If $n \geqslant 4$, then isometric mapping $\varphi_{\pi,g} \in \mathcal{I}_n$ preserves the Rayleigh quotient of every Boolean function in $n$ variables if and only if $\varphi_{\pi,g} \in \mathrm{Aut}\left(\mathrm{SB}^+(n)\right)$.

**Theorem 22** [14]. If $n \geqslant 4$, then isometric mapping $\varphi_{\pi,g} \in \mathcal{I}_n$ changes the sign of the Rayleigh quotient of every Boolean function in $n$ variables if and only if it is a bijection between $\mathrm{SB}^+(n)$ and $\mathrm{SB}^-(n)$.

In a scope of bent functions, the Rayleigh quotient characterizes the Hamming distance between a bent function and its dual. Indeed, let $f \in \mathcal{B}_n$, then

$$\mathrm{dist}(f, \widetilde{f}) = 2^{n-1} - \frac{S_f}{2^{n/2+1}}.$$

So from Theorem 21 we immediately have that general form of isometric mappings preserving the Hamming distance between every bent function and its dual is described by the extended orthogonal group $\overline{\mathcal{O}}_n$ (see Theorem 2).

## 8. Conclusion

In this paper, we have given a review of metrical properties of the set of bent functions and its subset of functions which coincide with their duals. The group of automorphisms and metrical complements of these sets are described. We also reviewed some general metrical properties of the set of self-dual bent functions and considered an iterative construction of bent functions. Some relevant open problems and hypothesis on bent functions were discussed.

An interesting question is the characterization of isometric mappings preserving bentness and self-duality, that are beyond the automorphisms of the set of all Boolean functions.

The solution of the problems, that were considered in this review, with regard to different generalizations of bent functions that is study of metrical properties and the duality as well as self-duality in this scope, is a goal worth pursuing.

REFERENCES

1. *Rothaus O. S.* On bent functions. J. Combin. Theory. Ser. A, 1976, vol. 20, no. 3, pp. 300–305.
2. *Tokareva N.* Bent Functions: Results and Applications to Cryptography. Acad. Press, Elsevier, 2015. 230 p.
3. *Carlet C. and Mesnager S.* Four decades of research on bent functions. Des. Codes Cryptogr., 2016, vol. 78, no. 1, pp. 5–50.
4. *Mesnager S.* Bent Functions: Fundamentals and Results. Berlin, Springer, 2016. 544 p.

5. *Kolomeec N.* The graph of minimal distances of bent functions and its properties. Des. Codes Cryptogr., 2017, vol. 85, no. 3, pp. 1–16.
6. *Janusz G. J.* Parametrization of self-dual codes by orthogonal matrices. Finite Fields Appl., 2007, vol. 13, no. 3, pp. 450–491.
7. *Dillon J.* Elementary Hadamard Difference Sets. PhD.dissertation, Univ. Maryland, College Park, 1974.
8. *Carlet C.* Boolean functions for cryptography and error correcting codes. Y. Crama and P. L. Hammer (eds.). Boolean Models and Methods in Mathematics, Computer Science, and Engineering. Cambridge, Cambridge University Press, 2010, pp. 257–397.
9. *Hou X.-D.* New constructions of bent functions. Proc. Intern. Conf. Combinatorics, Inform. Theory and Statistics. J. Combin. Inform. System Sci., 2000, vol. 25, no. 1–4, pp. 173–189.
10. *Cusick T. W. and Stănică P.* Cryptographic Boolean Functions and Applications. London, Acad. Press, 2017. 288 p.
11. *Tokareva N. N.* On the number of bent functions from iterative constructions: lower bounds. Adv. Math. Commun., 2011, vol. 5, no. 4, pp. 609–621.
12. *Tokareva N. N.* On decomposition of a Boolean function into sum of bent functions. Siberian Electronic Math. Reports, 2014, vol. 11, pp. 745–751.
13. *Tokareva N. N.* O razlozhenii dual'noy bent-funktsii v summu dvukh bent-funktsiy [On decomposition of a dual bent function into sum of two bent functions. Prikladnaya Diskretnaya Matematika, 2014, no. 4(26), pp. 59–61. (in Russian)
14. *Kutsenko A.* The group of automorphisms of the set of self-dual bent functions. Cryptogr. Commun., 2020, vol. 12, no. 5, pp. 881–898.
15. *Hou X.-D.* Classification of self dual quadratic bent functions. Des. Codes Cryptogr., 2012, vol. 63, no. 2, pp. 183–198.
16. *Carlet C., Danielson L. E., Parker M. G., and Solé P.* Self-dual bent functions. Int. J. Inform. Coding Theory, 2010, vol. 1, pp. 384–399.
17. *Feulner T., Sok L., Solé P. and Wassermann A.* Towards the classification of self-dual bent functions in eight variables. Des. Codes Cryptogr., 2013, vol. 68, no. 1, pp. 395–406.
18. *Hyun J. Y., Lee H., and Lee Y.* MacWilliams duality and Gleason-type theorem on self-dual bent functions. Des. Codes Cryptogr., 2012, vol. 63, no. 3, pp. 295–304.
19. *Mesnager S.* Several new infinite families of bent functions and their duals. IEEE Trans. Inf. Theory, 2014, vol. 60, no. 7, pp. 4397–4407.
20. *Rifà J. and Zinoviev V. A.* On binary quadratic symmetric bent and almost bent functions. 2019, arXiv:1211.5257v3.
21. *Mesnager S.* On constructions of bent functions from involutions. Proc. ISIT, 2016, pp. 110–114.
22. *Coulter R. and Mesnager S.* Bent functions from involutions over $\mathbb{F}_{2^n}$. IEEE Trans. Inf. Theory, 2018, vol. 64, no. 4, pp. 2979–2986.
23. *Luo G., Cao X., and Mesnager S.* Several new classes of self-dual bent functions derived from involutions. Cryptogr. Commun., 2019, vol. 11, no. 6, pp. 1261–1273.
24. *Sok L., Shi M., and Solé. P.* Classification and construction of quaternary self-dual bent functions. Cryptogr. Commun., 2018, vol. 10, no. 2, pp. 277–289.
25. *Kutsenko A.* Metrical properties of self-dual bent functions. Des. Codes Cryptogr., 2020, vol. 88, no. 1, pp. 201–222.
26. *Kutsenko A. V.* The Hamming distance spectrum between self-dual Maiorana-McFarland bent functions. J. Appl. Industr. Math., 2018, vol. 12, no. 1, pp. 112–125.
27. *McFarland R. L.* A family of difference sets in non-cyclic groups. J. Combin. Theory. Ser. A, 1973, vol. 15, no. 1, pp. 1–10.

28. *MacWilliams F. J. and Sloane N. J. A.* The Theory of Error-Correcting Codes. Amsterdam, New York, Oxford, North-Holland, 1983. 782 p.
29. *Canteaut A. and Charpin P.* Decomposing bent functions. IEEE Trans. Inform. Theory, 2003, vol. 49, no. 8, pp. 2004–2019.
30. *Preneel B., Van Leekwijck W., Van Linden L., et al.* Propagation characteristics of Boolean functions. Advances in Cryptology-EUROCRYPT, LNCS, 1990, vol. 473, pp. 161–173.
31. *Climent J.-J., Garcia F. J., and Requena V.* A construction of bent functions of $n+2$ variables from a bent function of $n$ variables and its cyclic shifts. Algebra, 2014, vol. 2014, Article ID 701298. 11 p.
32. *Canteaut A., Daum M., Dobertin H., and Leander G.* Finding nonnormal bent functions. Discrete Appl. Math., 2006, vol. 154, no. 2, pp. 202–218.
33. *Oblaukhov A. K.* Metric complements to subspaces in the Boolean Cube. J. Appl. Industr. Math., 2016, vol. 10, no. 3, pp. 397–403.
34. *Stănică P., Sasao T., and Butler J. T.* Distance duality on some classes of Boolean functions. J. Combin. Math. Combin. Computing, 2018, vol. 107, pp. 181–198.
35. *Oblaukhov A.* A lower bound on the size of the largest metrically regular subset of the Boolean cube. Cryptogr. Commun., 2019, vol. 11, no. 4, pp. 777–791.
36. *Tokareva N.* Duality between bent functions and affine functions. Discrete Math., 2012, vol. 312, no. 3, pp. 666–670.
37. *Markov A. A.* O preobrazovaniyakh, ne rasprostranyayushchikh iskazheniya [On transformations without error propagation]. Selected Works, vol. II: Theory of Algorithms and Constructive Mathematics. Mathematical Logic. Informatics and Related Topics, Moscow, MTsNMO Publ., 2003, pp. 70–93. (in Russian)
38. *Dempwolff U.* Automorphisms and equivalence of bent functions and of difference sets in elementary Abelian 2-groups. Commun. Algebra, 2006, vol. 34, no. 3, pp. 1077–1131.
39. *Tokareva N. N.* The group of automorphisms of the set of bent functions. Discrete Math. Appl., 2010, vol. 20, no. 5–6, pp. 655–664.
40. *Danielsen L. E., Parker M. G., and Solé P.* The Rayleigh quotient of bent functions. LNCS, 2009, vol. 5921, pp. 418–432.

UDC 519.7 DOI 10.17223/20710410/49/3

# ON METRIC COMPLEMENTS AND METRIC REGULARITY IN FINITE METRIC SPACES[1]

A. K. Oblaukhov

*Sobolev Institute of Mathematics, Novosibirsk, Russia,*
*Novosibirsk State University, Novosibirsk, Russia,*
*Laboratory of Cryptography JetBrains Research, Novosibirsk, Russia*

**E-mail:** oblaukhov@gmail.com

This review deals with the metric complements and metric regularity in the Boolean cube and in arbitrary finite metric spaces. Let $A$ be an arbitrary subset of a finite metric space $M$, and $\widehat{A}$ be the *metric complement* of $A$ — the set of all points of $M$ at the maximal possible distance from $A$. If the metric complement of the set $\widehat{A}$ coincides with $A$, then the set $A$ is called a *metrically regular set*. The problem of investigating metrically regular sets was posed by N. Tokareva in 2012 when studying metric properties of *bent functions*, which have important applications in cryptography and coding theory and are also one of the earliest examples of a metrically regular set. In this paper, main known problems and results concerning the metric regularity are overviewed, such as the problem of finding the largest and the smallest metrically regular sets, both in the general case and in the case of fixed covering radius, and the problem of obtaining metric complements and establishing metric regularity of linear codes. Results concerning metric regularity of partition sets of functions and Reed — Muller codes are presented.

**Keywords:** *metrically regular set, metric complement, covering radius, bent function, deep hole, Reed — Muller code, linear code.*

## 1. Introduction

The problem of investigating and classifying *metrically regular sets* was posed by N. Tokareva [1, 2] when studying metric properties of *bent functions* [3]. A Boolean function in even number of variables is called a *bent function* if it is at the maximal possible distance from the set of *affine functions*.

Bent functions have various applications in cryptography, coding theory and combinatorics [2, 4, 5]. In cryptography, bent functions are valued because of their outstanding nonlinearity, which helps to construct S-boxes for block ciphers with high resistance to linear cryptanalysis, and, as it turned out, good diffusion properties and high resistance to differential cryptanalysis [5]. Bent functions were also used in the construction of the stream cipher Grain, being a part of a nonlinear feedback shift register [2]. From the coding theory standpoint, bent functions form the set of points at the maximal possible distance from the Reed — Muller code of the first order $\mathcal{RM}(1, m)$ in even number of variables $m$. Bent functions are used to construct Kerdock codes, which are optimal and have large code distances (see more in [5]). Bent functions also have a number of representations

[1]The work was carried out under the state contract of the Sobolev Institute of Mathematics (project no. 0314-2019-0017) and supported by RFBR (projects no. 18-07-01394, 19-31-90093) and Laboratory of Cryptography JetBrains Research.

and relations to different combinatorial objects: Hadamard difference sets, block designs, etc. [2, 5].

However, many problems related to bent functions remain unsolved; in particular, the gap between the best known lower and upper bound on the number of bent functions is extremely large; currently known constructions of bent functions are rather scarse.

In 2010 [6], N. Tokareva has proved that, like bent functions are maximally distant from affine functions, affine functions are at the maximal possible distance from bent functions, thus establishing the *metric regularity* of both sets. Combined with the importance of bent functions in cryptography and coding theory, this arouses the interest in studying the property of metric regularity and in the classification of metrically regular sets.

This paper deals with the metrically regular sets in the Boolean cube and in arbitrary finite metric spaces. Published results concerning the topic, as well as some currently unpublished, are overviewed.

Section 2 provides necessary basic definitions, simple examples of metrically regular sets and some of their trivial properties. Section 3 describes the results of Stănică, Sasao and Butler [7] concerning metric complements and metric regularity of *partition sets of functions*. Section 4 deals with the problem of finding the smallest and the largest metrically regular sets, both in general and in the case of fixed distance between sets [8]. *Strongly metrically regular sets* are introduced in Section 5 as a subclass of metrically regular sets. These allow one to obtain iterative constructions of metrically regular sets and get an estimate on how big the largest metrically regular set with fixed covering radius can be [9]. Section 6 touches upon the problem of describing metric complements and establishing metric regularity of linear codes. General results are presented, and the metric regularity of several families of Reed — Muller codes is established [10, 11].

## 2. Preliminaries

### 2.1. Definitions

Let $M$ be a finite discrete metric space with a metric $d(\cdot,\cdot)$, which admits values from a set $D$. From now on, every space mentioned in the paper will be a finite discrete metric space. Let $X \subseteq M$ be an arbitrary subset of the space (in this paper, whenever the symbol "$\subset$" is used, it will imply a nonempty proper subset) and $y \in M$ be an arbitrary point. The distance $\mathrm{d}(y, X)$ from the point $y$ to the set $X$ is equal to $\min\limits_{x\in X} \mathrm{d}(y, x)$. The *covering radius* of the set $X$ is defined as follows:

$$\rho(X) = \max_{z\in M} \mathrm{d}(z, X).$$

A set $X$ with the covering radius $r$ is also sometimes called a *covering code* [12] of radius $r$.

Consider the following set

$$\{y \in M : \mathrm{d}(y, X) = \rho(X)\}$$

of all vectors at the maximal possible distance from the set $X$. This set is called the *metric complement* [10] of $X$ and is denoted by $\widehat{X}$. If $\widehat{\widehat{X}} = X$, the set $X$ is said to be *metrically regular* [1].

Note that metrically regular sets always come in pairs, i.e. if $A$ is a metrically regular set, its metric complement $\widehat{A}$ is also a metrically regular set. In this paper, a pair consisting of a metrically regular set $A$ and its metric complement $B = \widehat{A}$ will sometimes be referred to as "a pair of metrically regular sets $A$, $B$".

Throughout the paper, we will mostly consider the metric space $\mathbb{F}_2^n$ of binary vectors of length $n$ equipped with the Hamming metric. The *Hamming distance* $\mathrm{d_H}(\cdot,\cdot)$ between

two binary vectors is defined as the number of coordinates in which these vectors differ, while $\mathrm{wt}(\cdot)$ denotes the *Hamming weight* of a vector, i.e., the number of nonzero values it contains. Since $\mathbb{F}_2$ is a field, $\mathbb{F}_2^n$ is also considered as a vector space with the plus sign "+" denoting addition of vectors modulo two. A *Boolean function* in $m$ variables is an arbitrary mapping from $\mathbb{F}_2^m$ to $\mathbb{F}_2$.

### 2.2. Examples and basic results

Let us consider some simple examples of metric complements and metrically regular sets in the space $\mathbb{F}_2^n$.

1) Let $X = \{x\}$ be the set consisting of one binary vector. It has covering radius $n$ and its metric complement is the set $\widehat{X} = \{x + \mathbf{1}\}$, consisting only of the opposite vector (here $\mathbf{1}$ is the all-ones vector). It follows that $\widehat{\widehat{X}} = X$, so $X$ is a metrically regular set.
2) Consider a ball of radius $r$ centered at $x$, i.e., $X = \{y \in \mathbb{F}_2^n : \mathrm{d}(x, y) \leqslant r\}$. Then the vector $x + \mathbf{1}$ will be at the distance $n - r$ from the set $X$, while any other vector will be at a smaller distance. Therefore, the covering radius of $X$ is equal to $n - r$ and its metric complement is the set $\widehat{X} = \{x + \mathbf{1}\}$. Then $\widehat{\widehat{X}} = \{x\}$, which shows us that, unless $r = 0$, the ball of radius $r$ is not a metrically regular set.

For other examples of metric complements and metrically regular sets the reader is referred to [8 – 10].

Let us return to an arbitrary metric space $M$ with a metric admitting values from a set $D$ and present some basic results concerning metric regularity.

An *automorphism* of a set $X \subseteq M$ is an isometric mapping from $M$ into $M$ which maps $X$ into itself. The following result [10] is straightforward from the definition of metric regularity, and is also described in [6, 1] for affine/bent functions.

**Theorem 1** [10]. Let $X \subset M$ be a metrically regular set. Then sets of automorphisms of $X$ and $\widehat{X}$ coincide: $\mathrm{Aut}(X) = \mathrm{Aut}(\widehat{X})$.

As we could see from examples, not every set is metrically regular, which means that we can apply the procedure of taking metric complement more than twice and obtain new sets. It has been proven [10] that this process stabilizes for any set after not more than $|D| - 1$ repetitions.

**Proposition 1** [10]. Let $X$ be an arbitrary subset of $M$. Let us denote $X_0 = X$, $X_{k+1} = \widehat{X}_k$ for $k \geqslant 0$. Then there exists a number $N \leqslant |D| - 1$ such that $X_n$ is a metrically regular set for any $n \geqslant N$.

Using this proposition, we can, for example, split the set $2^M$ of all subsets of $M$ into equivalence classes, and call two sets $X$,$Y \subseteq M$ equivalent if and only if the pair of metrically regular sets $A, A^*$, which we obtain from the set $X$ by repeatedly obtaining metric complement as in Proposition 1, coincides with the pair of metrically regular sets $B, B^*$ which we obtain from the set $Y$. How would the equivalence classes look? The description has not yet been given.

Proposition 1 is also useful when conducting experiments with metrically regular sets using computers.

## 3. Partition sets of functions

In [7], authors introduce the notion of *partition sets of functions* and study their metric complements and metric regularity.

A set $\mathcal{S}$ of Boolean functions in $m$ variables is said to be a *partition set* with respect to a partition $\mathcal{U}$ of the set $\mathbb{F}_2^m$, if the elements in the same block of $\mathcal{U}$ all map to 0 or all map to 1, and all combinations of assignments to the blocks are included in $\mathcal{S}$. Partition set functions include, for example, symmetric functions, rotation symmetric functions, self-anti-dual-functions and linear structure functions.

The following theorem presents the main result of [7], describing the covering radius and the metric complement of a partition set of functions.

**Theorem 2** [7]. Consider a partition set of functions $\mathcal{S}$, and let us denote the covering radius of $\mathcal{S}$ as $\rho_{\mathcal{S}}$. Let $N_{\mathcal{S}}$ be the number of Boolean functions at distance $\rho_{\mathcal{S}}$ from $\mathcal{S}$. Then,

$$\rho_{\mathcal{S}} = \sum_{i=1}^{l} \lfloor k_i/2 \rfloor \quad \text{and} \quad N_{\mathcal{S}} = \prod_{i=1}^{l} \frac{1}{2 - k_i \bmod 2} \left( \binom{k_i}{\lfloor k_i/2 \rfloor} + \binom{k_i}{\lceil k_i/2 \rceil} \right),$$

where $k_i$ is the cardinality of the $i$-th block of the $l$ blocks in partition $\mathcal{U}$.

The proof of the theorem is constructive and gives an explicit description of the metric complement $\widehat{\mathcal{S}}$. From this description, the equality $\widehat{\widehat{\mathcal{S}}} = \mathcal{S}$ is trivially established, showing that all partition sets of functions are metrically regular.

The authors then proceed to investigate special cases of partition sets of functions, namely, *symmetric* and *rotation symmetric* functions. They calculate covering radii for both of these sets, give characterization for the set of maximally asymmetric functions (the metric complement of the set of symmetric functions) and calculate the number of such functions. They also study the weight distribution of maximally asymmetric functions, as well as their algebraic degrees, and provide a classification of all functions with respect to the distance from the set of symmetric functions. For details, the reader is referred to [7].

## 4. Largest and smallest metrically regular sets

Let us return to affine and bent functions. Since the gap between the best known upper and lower bounds on the size of the set of bent functions is so large, it is interesting to investigate possible cardinalities of metrically regular sets, particularly, the extreme cardinalities, in an attempt to improve known bounds. The paper [8] focuses on the problem of finding the largest and the smallest metrically regular sets.

### 4.1. General problem

In the Boolean cube $\mathbb{F}_2^n$ with the Hamming distance, any smallest metrically regular set has cardinality 1, as can be seen from the simplest example $X = \{x\}$, $x \in \mathbb{F}_2^n$. For the largest metrically regular set the solution is not so trivial. The following theorem reduces the general problem to a special case.

**Theorem 3** [8]. Let $A, B \subset \mathbb{F}_2^n$ be a pair of metrically regular sets, i.e., $A = \widehat{B}$, $B = \widehat{A}$. Then there exists a pair of metrically regular sets $A^*$, $B^*$ at distance 1 from each other such that either $A \subseteq A^*$, $B \subseteq B^*$, or both $A$,$B \subseteq A^*$.

The Theorem 3 tells us that for each metrically regular set in the Boolean cube there exists a metrically regular superset with the covering radius of 1. Therefore, the covering radius of the largest metrically regular set in the Boolean cube is equal to 1. Since for any set $A$ with $\rho(A) = 1$ it holds $A \cup \widehat{A} = \mathbb{F}_2^n$, the largest metrically regular set is the metric (and ordinary) complement of the smallest metrically regular set with the covering radius equal to 1.

The problem is reduced further by the following fact.

**Proposition 2** [8]. If $C \subseteq \mathbb{F}_2^n$ is a minimal covering code of radius 1, then $C$ is metrically regular.

It follows from the Proposition 2 that any smallest covering code of radius 1 is also a smallest metrically regular set with the covering radius 1. Combined with Theorem 3, this shows that the problem of finding the largest metrically regular set is equivalent to the problem of finding the smallest covering code of radius 1. This is an open problem of coding theory [12] and is solved mostly for particular cases and small dimensions.

Proposition 2 is conjectured to hold true for larger values of the covering radius, however, this has not been proved yet.

**Conjecture 1** [8]. If $C \subseteq \mathbb{F}_2^n$ is a covering code of radius $r$ of minimal size, then $C$ is metrically regular.

The conjecture was computationally checked [8] for several minimal covering codes with $n = 2r+3, 2r+4$, where $r$ equals 2 or 3. Constructions of these codes can be found in [13, 14].

### 4.2. Fixed distances

As we see from the previous subsection, the general problems of finding the largest and the smallest metrically regular sets are reduced to the cases when the covering radius is trivial (equal to either 1 or $n$). However, the set $\mathcal{B}_m$ of bent functions in $m$ variables has the covering radius $2^{m-1} - 2^{m/2-1}$. In [8], the sizes of the sets at a fixed distance $r$ from each other are considered. Theses sizes are estimated nondirectly, through estimating the size of the union of two metrically regular sets, maximally distant one from another. Let us return to the general finite metric space $M$ with a metric $d(\cdot, \cdot)$ admitting values from a set $D$. Then, the following bound holds.

**Theorem 4** [8]. Let $A$,$B \subseteq M$ be a pair of metrically regular sets at distance $r \in D$ from each other, and let $C_k$ be the size of the largest sphere of radius $k \in D$ in $M$. Then

$$|A| + |B| \geqslant \frac{2|M|}{1 + \sum\limits_{\substack{k \in D \\ k<r}} C_k}.$$

This bound is very similar to the sphere-packing bound on the size of a code, well-known in the coding theory. In the case when the space $M$ is $\mathbb{F}_2^n$ with the Hamming metric, the bound becomes:

**Corollary 1.** Let $A$,$B \subseteq \mathbb{F}_2^n$ be a pair of metrically regular sets at distance $r$ from each other. Then

$$|A| + |B| \geqslant \frac{2^{n+1}}{1 + \sum\limits_{k=0}^{r-1} \binom{n}{k}}.$$

## 5. Strongly metrically regular sets

### 5.1. Preliminaries

Metrically regular sets are defined by their outstanding metric properties, but a lot of them possess even more regularity. In order to investigate largest and smallest metrically regular sets further, the notion of a *strongly metrically regular* set was introduced in [9].

Let $A \subset \mathbb{F}_2^n$ be a set with the covering radius $r$. The set $A$ is called *strongly metrically regular*, if for any vector $x \in \mathbb{F}_2^n$ it holds

$$\mathrm{d}(x, A) + \mathrm{d}(x, \widehat{A}) = r.$$

In other words, any vector of the Boolean cube belongs to some shortest path from the set $A$ to the set $\widehat{A}$. It is clear from the definition that any strongly metrically regular set is metrically regular.

The following pair of metrically regular sets gives us a simple example: $A = \{\mathbf{0}\}, \widehat{A} = \{\mathbf{1}\}$. Any vector $x \in \mathbb{F}_2^n$ with the Hamming weight $k$ is at distance $k$ from the set $A$ and at distance $(n-k)$ from the set $\widehat{A}$, so the sum of both distances is equal to $n$, which is the covering radius of these sets.

But not all metrically regular sets are strongly metrically regular. One of the problems of the International Cryptographic Olympiad NSUCRYPTO 2016 [15] was to find a metrically regular set which is not strongly metrically regular (or prove that such set does not exist), and several contestants managed to find a solution. The smallest known example of such a set is contained in the Boolean cube of dimension 7.

Let $A$ be an arbitrary subset of the Boolean cube $\mathbb{F}_2^n$. The *layer representation* of $\mathbb{F}_2^n$ with respect to the set $A$ is the sequence of layers defined as follows:

$$A_k = \{x \in \mathbb{F}_2^n : \mathrm{d}(x, A) = k\},\ k = 0, 1, \ldots, r,$$

where $r$ is the covering radius of $A$. Using layer representation, strongly metrically regular sets can alternatively be defined as follows:

**Proposition 3** [9]. Set $A$ is strongly metrically regular if and only if for any $k$ from 0 to $r$ it holds $A_k = \widehat{A}_{r-k}$, where $r$ is the covering radius of both sets.

It is easy to see that completely regular codes [16] are strongly metrically regular. The converse is not true: an example of a strongly metrically regular set which is not a completely regular code is the set $A = \{(000), (011), (111)\}$ in $\mathbb{F}_2^3$.

### 5.2. Iterative constructions

In [9], several iterative constructions of strongly metrically regular sets are obtained.

**Theorem 5** [9]. Let $A$ be a strongly metrically regular set with the covering radius $r$. Then $C = A \cup \widehat{A}$ is also a strongly metrically regular set.

Then this theorem is generalized to obtain more iterative constructions of strongly metrically regular sets.

**Theorem 6.** Let $A$ be a strongly metrically regular set with the covering radius $r > 0$ (case $r = 0$ is trivial). Let $i_1, \ldots, i_s$ be a sequence of indices satisfying $0 \leqslant i_1 < i_2 < \ldots < i_{s-1} < i_s \leqslant r$. Then the union $C = \bigcup\limits_{k=1}^{s} A_{i_k}$ is a strongly metrically regular set if and only if there exists a number $\rho > 0$ such that all the following conditions are satisfied:

1) for any $k \in \{1, \ldots, s-1\}$ the distance $(i_{k+1} - i_k)$ is equal to 1, $2\rho$ or $2\rho + 1$;
2) for any $k \in \{2, \ldots, s-1\}$ at least one of the distances $(i_{k+1} - i_k), (i_k - i_{k-1})$ is greater than 1;
3) $i_1$ is either $\rho$ or 0, and if $i_1 = 0$, then $i_2 - i_1 = 2\rho$ or $2\rho + 1$ if $i_2$ exists;
4) $i_s$ is either $r - \rho$ or $r$, and if $i_s = r$, then $i_s - i_{s-1} = 2\rho$ or $2\rho + 1$ if $i_{s-1}$ exists;

The number $\rho$ is the covering radius of $C$.

Theorem 6 allows one to construct many new strongly metrically regular sets with smaller covering radii given a strongly metrically regular set with the covering radius $r$. For example, consider a strongly metrically regular set with the covering radius 20. Then, if we take the union of layers with indices $\{2, 3, 7, 12, 16, 20\}$, it will be a strongly metrically regular set with the covering radius 2 and its metric complement will consist of layers with indices $\{0, 5, 9, 10, 14, 18\}$.

The number of strongly metrically regular sets with the covering radius $r$ which can be constructed using Theorem 6 is also calculated.

**Theorem 7** [9]. Let $A$ be a strongly metrically regular set with the covering radius $r > 0$. Then the number $G_\rho(r)$ of different strongly metrically regular sets with covering radius $\rho$ that can be obtained by applying Theorem 6 to the set $A$ can be calculated using the following recurrent formulas:

$$G_\rho(r) = \begin{cases} G_\rho(r-\rho) + G_\rho(r-\rho-1), & \text{when } r > \rho, \\ 2, & \text{when } r = \rho, \\ 0, & \text{when } 0 \leqslant r < \rho. \end{cases}$$

### 5.3. Special constructions and lower bounds

Utilizing Theorem 6 and other considerations, two families of "large" strongly metrically regular sets $\{Y_n^r\}$, $\{Z_n^r\}$ for $n \geqslant 2r$, $r \geqslant 1$ are constructed in [9]. Here, $Y_n^r, Z_n^r \subseteq \mathbb{F}_2^n$ and $\rho(Y_n^r) = \rho(Z_n^r) = r$. Sets from these families asymptotically cover a large part of the Boolean cube:

$$|Y_n^r| \overset{n\to\infty}{\sim} \frac{2}{2r+1}2^n, \quad |Z_n^r| = 2^{n-2r}\binom{2r}{r} \overset{r\to\infty}{\sim} \frac{1}{\sqrt{\pi r}}2^n.$$

The lower bound on the sizes of sets from the family $\{Y_n^r\}$ is obtained, which results in the following lower bound on the size of the largest metrically regular set for fixed covering radius.

**Theorem 8.** Let $A$ be the largest metrically regular set with the covering radius $r$ in the Boolean cube of dimension $n$ ($n \geqslant 2r$), and let $\rho$ be the remainder of $n+1$ divided by $2r+1$. Then

$$|A| \geqslant \max\left\{2^n\left(\frac{2}{2r+1} - \frac{2}{\sqrt{n-\rho+1}}\right), 2^{n-2r}\binom{2r}{r}\right\}.$$

Construction of the family of strongly metrically regular sets $\{Y_n^r\}$ allows one to obtain metrically regular sets with the covering radius $r$ that cover roughly the fraction $\dfrac{2}{2r+1}$ of the whole Boolean cube when $n$ is big enough, while the family $\{Z_n^r\}$ contains metrically regular sets with the covering radius $r$ that cover roughly the fraction $\dfrac{1}{\sqrt{\pi r}}$ of the Boolean cube for large values of $r$.

## 6. Metric complements and metric regularity of linear codes

### 6.1. General results

The papers [10, 11] touch upon the topic of metric complements of linear codes in the Boolean cube. First, let us formulate some basic results.

**Proposition 4.** Let $L \subseteq \mathbb{F}_2^n$ be a linear code. Then the metric complement of $L$ is the union of cosets of $L$.

This result follows directly from the equality $\mathrm{d_H}(x, y) = \mathrm{wt}(x + y)$ and the linearity of the code. The following bound is also a simple and well-known result.

**Proposition 5.** Let $L \subseteq \mathbb{F}_2^n$ be a linear code of dimension $k$. Then $\rho(L) \leqslant n - k$.

The paper [10] describes sufficient and necessary conditions on an arbitrary linear code $L$ to attain this bound, as well as some sufficient conditions for $\rho(L) = n - k - 1$ or $\rho(L) = n - k - 2$. Both of these results also present explicit form of the metric complement of the linear code in question, and in the case when $\rho(L) = n - k$, the code $L$ is found to be metrically regular.

The following characterization of the second metric complement using the first is also presented in [10, 1].

**Proposition 6.** Let $L \subseteq \mathbb{F}_2^n$ be a linear code. Then $\rho(\widehat{L}) = \rho(L)$ and a vector $x$ is in $\widehat{\widehat{L}}$ if and only if $x + \widehat{L} = \widehat{L}$.

**Corollary 2.** Let $L \subseteq \mathbb{F}_2^n$ be a linear code. Assume that $\widehat{L}$ is an affine subspace, i.e., $\widehat{L} = a + L_1$ for some linear code $L_1$. Then $\widehat{\widehat{L}} = L_1$.

### 6.2. Sets of affine/bent functions

Let us remember that the notion of a metrically regular set and the problem of investigating and classifying metrically regular sets was first posed by N. Tokareva in [1] when studying metric properties of bent functions, particularly, the duality between bent functions and affine functions.

A Boolean function in even number $m$ of variables is called a *bent function*, if it is at the maximal possible distance from the set of affine functions $\mathcal{A}_m$. If we denote the set of bent functions as $\mathcal{B}_m$, then we have, by definition, $\mathcal{B}_m = \widehat{\mathcal{A}}_m$.

Despite the fact that all characterizations of the set of bent functions that are currently known are rather ineffective when it comes to counting and constructing bent-functions, it turned out that these characterizations are enough to establish metric regularity of the set of affine/bent functions.

It follows from Proposition 6 that a linear code is metrically regular if and only if no vectors other that those from the code keep its metric complement stable under addition. This property of linear codes was used in [6, 1] to establish that the set of affine functions is the metric complement of the set of bent functions: N. Tokareva has shown that, for any non-affine function $f$, there exists a bent function $g$ (from the Maiorana — McFarland class of bent functions) such that $f + g$ is not a bent function. Thus, the following holds.

**Theorem 9.** Sets of affine functions $\mathcal{A}_m$ and bent functions $\mathcal{B}_m$ are metrically regular.

A. Kutsenko studied metric properties of two subclasses of bent functions called *self-dual* and *anti-self-dual* bent functions. In [17], he shows that the set of self-dual bent functions is the metric complement of the set of anti-self-dual bent functions and vice versa, thus establishing the metric regularity of both of these sets. Other metric properties of bent functions (e.g. the graph of minimal distances between bent functions) were also studied by N. Kolomeec in [18 – 21].

### 6.3. Reed — Muller codes

Let $\mathcal{F}^m$ be the set of all Boolean functions in $m$ variables. The Reed — Muller code of order $k$ in $m$ variables is defined as follows:

$$\mathcal{RM}(k, m) = \{f \in \mathcal{F}^m : \deg(f) \leqslant k\},$$

where $\deg(\cdot)$ denotes the degree of the *algebraic normal form* [2] of the function. These codes may also be represented as sets of *value vectors* of corresponding functions: binary vectors of length $2^m$, containing values which a function assumes on all vectors of $\mathbb{F}_2^m$, listed in some fixed order. Distances between functions can therefore be defined as distances between their value vectors.

The Reed — Muller code of order 1 is, by definition, the set of affine functions, which is, in the case of even number of variables $m$, metrically regular (as is its metric complement — the set of bent functions). Does this hold for other codes from this family? In [11], this metric property for other Reed — Muller codes is being investigated.

In [22], E. Berlekamp and N. Welch presented a partition of all cosets of the $\mathcal{RM}(1,5)$ code into 48 classes with respect to the EA-equivalence (extended affine equivalence), providing a representative for each class. Then they obtained weight distributions for each class of cosets. This weight distribution allows one to explicitly describe the metric complement of the code by selecting classes with the largest minimal weigth. Proposition 6 is then used to establish the metric regularity of $\mathcal{RM}(1,5)$ in [11]. It is shown that, for any equivalence class of cosets (other than the $\mathcal{RM}(1,5)$ itself), adding a function from that class to some function from the metric complement $\widehat{\mathcal{RM}}(1,5)$ yields a function outside of the metric complement, leading to the following

**Theorem 10.** The code $\mathcal{RM}(1,5)$ is metrically regular.

Reed — Muller codes of orders 0, $m$ and $m-1$ coincide with the repetition code, the whole space, and the even weight code respectively. It is trivial that all of them are metrically regular. Metric regularity of the Reed — Muller code of order $m-2$ is also easy to establish as follows [11].

The Reed — Muller code of order $m-2$ has covering radius 2 [12]. By definition, it consists of all Boolean functions of degree at most $m-2$. Since all functions of degree $m$ have odd weights, and all functions of smaller degree have even weights, functions of degree $m$ are at distance 1 from $\mathcal{RM}(m-2,m)$, while functions of degree $m-1$ are at distance 2, and therefore

$$\widehat{\mathcal{RM}}(m-2,m) = \mathcal{RM}(m-1,m) \setminus \mathcal{RM}(m-2,m).$$

Since $\mathcal{RM}(m-2,m)$ is linear, $\rho(\widehat{\mathcal{RM}}(m-2,m)) = \rho(\mathcal{RM}(m-2,m)) = 2$ and thus functions of degree $m$ are at distance 1 from $\widehat{\mathcal{RM}}(m-2,m)$. It follows that $\widehat{\widehat{\mathcal{RM}}}(m-2,m) = \mathcal{RM}(m-2,m)$ and therefore the following holds:

**Theorem 11.** Codes $\mathcal{RM}(k,m)$ for $k \geqslant m-2$ are metrically regular.

Codes of order $m-3$ are harder to handle. In 1979, A. M. McLoughlin [23] has proved that

$$\rho(\mathcal{RM}(m-3,m)) = \begin{cases} m+1, & \text{if } m \text{ is odd,} \\ m+2, & \text{if } m \text{ is even.} \end{cases}$$

This result is reestablished by G. Cohen et al. in [12] using a method of syndrome matrices, different from the method in [23]. This method allows the author of [11] not only to obtain the covering radius of the Reed — Muller code of order $m-3$, but also to describe the metric complement of this code. As with the covering radius, the cases of even and odd $m$ are distinct.

In the case of even number $m$ of variables, the metric complement can be described as follows:

$$\widehat{\mathcal{RM}}(m-3,m) = \bigcup_{g \in G} (g + \mathcal{RM}(m-3,m)),$$

where

$$G = \{g : \operatorname{supp}(g) = \{\mathbf{0}, \mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m, \mathbf{x}_1 + \ldots + \mathbf{x}_m\}, \mathbf{x}_1, \ldots, \mathbf{x}_m \text{ are linearly independent}\},$$

while, for $m$ odd, the description is as follows:

$$\widehat{\mathcal{RM}}(m-3,m) = \bigcup_{g \in G_1 \cup G_2} (g + \mathcal{RM}(m-3,m)),$$

$$G_1 = \{g : \operatorname{supp}(g) = \{\mathbf{0}, \mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m\}, \mathbf{x}_1, \ldots, \mathbf{x}_m \text{ are linearly independent}\},$$

$$G_2 = \{g : \operatorname{supp}(g) = \{\mathbf{0}, \mathbf{x}_1, \ldots, \mathbf{x}_{m-1}, \mathbf{x}_1 + \ldots + \mathbf{x}_{m-1}\}, \mathbf{x}_1, \ldots, \mathbf{x}_{m-1} \text{ are linearly independent}\}.$$

Then, the metric regularity of $\mathcal{RM}(m-3,m)$ is proved by establishing that no functions other that those contained in $\mathcal{RM}(m-3,m)$ preserve the metric complement under addition (once again utilizing Proposition 6).

The author then considers the code $\mathcal{RM}(2,6)$. Using a proper ordering of the values in the value vectors of functions, this code can be presented in the following manner:

$$\mathcal{RM}(2,6) = \{(\mathbf{u}, \mathbf{u}+\mathbf{v}) : \mathbf{u} \in \mathcal{RM}(2,5), \mathbf{v} \in \mathcal{RM}(1,5)\}.$$

Since both $\mathcal{RM}(2,5)$ and $\mathcal{RM}(1,5)$ were shown to be metrically regular, this constructions is useful and allows the author to establish the metric regularity of the code $\mathcal{RM}(2,6)$ as well. The proof of this result heavily relies on the fact that $\mathcal{RM}(2,6)$ attains the upper bound on the covering radius provided by the $(\mathbf{u}, \mathbf{u}+\mathbf{v})$ construction, i.e., $\rho(\mathcal{RM}(2,6)) = \rho(\mathcal{RM}(2,5)) + \rho(\mathcal{RM}(1,5))$ [24].

Thus, the metric regularity of the codes $\mathcal{RM}(1,5)$, $\mathcal{RM}(2,6)$ and of the codes $\mathcal{RM}(k,m)$ for $k \geqslant m-3$ has been established. Factoring in the result by N. Tokareva [6], which proves the metric regularity of $\mathcal{RM}(1,m)$ for even $m$, this covers all infinite families of Reed — Muller codes with known covering radius. The only other Reed — Muller codes with known covering radius, metric regularity of which has not been yet established, are $\mathcal{RM}(1,7)$ [25, 26] and $\mathcal{RM}(2,7)$ [27]. Given these results, the following conjecture is formulated [11].

**Conjecture 2.** All Reed — Muller codes $\mathcal{RM}(k,m)$ are metrically regular.

## 7. Conclusion

In the paper, the main published results concerning metric complements and metric regularity are presented. Metric regularity of partition sets of functions is established. General problem of finding smallest metrically regular sets is found to be trivial, while finding the largest is shown to be as hard as finding the smallest covering code of radius 1. For fixed covering radius, a lower bounds on the sum of sizes of metrically regular sets constituting a pair is obtained. Using the notion of strongly metrically regular set, iterative constructions of metrically regular sets are described and the number of sets which can be obtained using these constructions is calculated. Two families of "large" (relative to the size of $\mathbb{F}_2^n$) metrically regular sets with fixed covering radius are constructed, giving the idea of how big the largest metrically regular sets can be. Characterizations of the first and the second metric complements of linear codes are given. Metric regularity of the Reed — Muller codes $\mathcal{RM}(1,m)$ for $m$ even, $\mathcal{RM}(k,m)$ for $k=0$, $k \geqslant m-3$ and of the codes $\mathcal{RM}(1,5)$ and $\mathcal{RM}(2,6)$ is established.

REFERENCES

1. *Tokareva N.* Duality between bent functions and affine functions. Discrete Mathematics, 2012, vol. 312, no. 3, pp. 666–670.
2. *Tokareva N.* Bent Functions: Results and Applications to Cryptography. Academic Press, 2015. 220 p.
3. *Rothaus O. S.* On "bent" functions. J. Combin. Theory. Ser. A, 1976, vol. 20, no. 3, pp. 300–305.
4. *Cusick T. W. and Stănică P.* Cryptographic Boolean Functions and Applications. Academic Press, 2017. 288 p.
5. *Mesnager S.* Bent Functions: Fundamentals and Results. Springer International Publishing, 2016. 544 p.
6. *Tokareva N. N.* The group of automorphisms of the set of bent functions. Discrete Math. Appl., 2010, vol. 20, no. 5–6, pp. 655–664.

7. *Stănică P., Sasao T., and Butler J. T.* Distance duality on some classes of Boolean functions. J. Combin. Math. Combin. Computing, 2018, vol. 107, pp. 181–198.
8. *Oblaukhov A. K.* Maximal metrically regular sets. Siberian Electronic Math. Reports, 2018, vol. 15, pp. 1842–1849.
9. *Oblaukhov A.* A lower bound on the size of the largest metrically regular subset of the Boolean cube. Cryptogr. Commun., 2019, vol. 11, no. 4, pp. 777–791.
10. *Oblaukhov A. K.* Metric complements to subspaces in the Boolean cube. J. Appl. Industr. Math., 2016, vol. 10, no. 3, pp. 397–403.
11. *Oblaukhov A.* `https://arxiv.org/abs/1912.10811` — On metric regularity of Reed — Muller codes, 2020.
12. *Cohen G., Honkala I., Litsyn S., and Lobstein A.* Covering Codes. Elsevier, 1997, vol. 54.
13. *Cohen G., Lobstein A., and Sloane N.* Further results on the covering radius of codes. IEEE Trans. Inform. Theory, 1986, vol. 32, no. 5, pp. 680–694.
14. *Graham R. L. and Sloane N.* On the covering radius of codes. IEEE Trans. Inform. Theory, 1985, vol. 31, no. 3, pp. 385–401.
15. *Tokareva N., Gorodilova A., Agievich S., et al.* Mathematical methods in solutions of the problems presented at the Third International Students' Olympiad in Cryptography. Prikladnaya Diskretnaya Matematika, 2018, no. 40, pp. 34–58.
16. *Neumaier A.* Completely regular codes. Discrete Math., 1992, vol. 106, pp. 353–360.
17. *Kutsenko A.* Metrical properties of self-dual bent functions. Designs, Codes Cryptography, 2020, vol. 88, no. 1, pp. 201–222.
18. *Kolomeec N. A. and Pavlov A. V.* Svoystva bent-funktsiy, nakhodyashchikhsya na minimal'nom rasstoyanii drug ot druga [Properties of bent functions which are at minimal distance from each other]. Prikladnaya Diskretnaya Matematika, 2009, no. 4(6), pp. 5–20. (in Russian)
19. *Kolomeets N. A.* Enumeration of the bent functions of least deviation from a quadratic bent function. J. Appl. Industr. Math., 2012, vol. 6, no. 3, pp. 306–317.
20. *Kolomeec N. A.* Verkhnyaya otsenka chisla bent-funktsiy na rasstoyanii $2^k$ ot proizvol'noy bent-funktsii ot 2k peremennykh [Upper bound on the number of bent functions which are at distance $2^k$ from an arbitrary bent function]. Prikladnaya Diskretnaya Matematika, 2014, no. 3(25), pp. 28–39. (in Russian)
21. *Kolomeec N.* The graph of minimal distances of bent functions and its properties. Designs, Codes Cryptography, 2017, vol. 85, no. 3, pp. 395–410.
22. *Berlekamp E. and Welch N.* Weight distributions of the cosets of the $(32, 6)$ Reed — Muller code. IEEE Trans. Inform. Theory, 1972, vol. 18, no. 1, pp. 203–207.
23. *McLoughlin A. M.* The covering radius of the $(m-3)$-rd order Reed — Muller codes and a lower bound on the $(m-4)$-th order Reed Muller codes. SIAM J. Appl. Math., 1979, vol. 37, no. 2, pp. 419–422.
24. *Schatz J.* The second order Reed — Muller code of length 64 has covering radius 18. IEEE Trans. Inform. Theory, 1981, vol. 17, no. 4, pp. 529–530.
25. *Mykkeltveit J.* The covering radius of the $(128, 8)$ Reed — Muller code is 56. IEEE Trans. Inform. Theory, 1980, vol. 26, no. 3, pp. 359–362.
26. *Hou X. D.* Covering radius of the Reed — Muller code $R(1, 7)$ — a simpler proof. J. Combin. Theory. Ser. A, 1996, vol. 74, no. 2, pp. 337–341.
27. *Wang Q.* The covering radius of the Reed — Muller code $RM(2, 7)$ is 40. Discrete Math., 2019, vol. 342, no. 12, Article 111625.

# МАТЕМАТИЧЕСКИЕ МЕТОДЫ КРИПТОГРАФИИ

УДК 519.7

## О ВОЗМОЖНОСТИ ПРИМЕНЕНИЯ СВЁРТОЧНЫХ НЕЙРОННЫХ СЕТЕЙ К ПОСТРОЕНИЮ УНИВЕРСАЛЬНЫХ АТАК НА ИТЕРАТИВНЫЕ БЛОЧНЫЕ ШИФРЫ

А. А. Перов*, А. И. Пестунов**

**Московский политехнический университет, г. Москва, Россия*

***Новосибирский государственный университет экономики и управления "НИНХ", г. Новосибирск, Россия*

Исследуется возможность применения свёрточных нейронных сетей к задаче анализа стойкости итеративных блочных шифров. Предлагается новый подход к построению атак-различителей на основе свёрточной нейронной сети, обученной различать графические эквиваленты шифртекстов, полученных в режиме шифрования CTR (счётчика) после разного числа раундов, в том числе после такого, которое обеспечивает удовлетворительные статистические свойства шифртекста. По аналогии со статистическими тестами, предложенный подход позволяет создавать различители без необходимости проведения аналитического исследования каждого шифра, что даёт возможность строить универсальные различители сразу для серии шифров. Предлагается несколько схем построения универсальных атак-различителей, которые, как демонстрируется экспериментально, в ряде случаев позволяют выявлять отклонения от случайности на меньших выборках и при большем числе раундов, чем ранее известные статистические тесты.

**Ключевые слова:** *блочный шифр, машинное обучение, нейронная сеть, статистический анализ, атака-различитель, криптоанализ.*

DOI 10.17223/20710410/49/4

## ON POSSIBILITY OF USING CONVOLUTIONAL NEURAL NETWORKS FOR CREATING UNIVERSAL ATTACKS ON ITERATIVE BLOCK CIPHERS

A. A. Perov*, A. I. Pestunov**

**Moscow Polytechnic University, Moscow, Russia*

***Novosibirsk State University of Economics and Management, Novosibirsk, Russia*

**E-mail:** perov_artem@inbox.ru, pestunov@gmail.com

The paper explores possibility of applying convolutional neural networks to the security analysis of iterative block ciphers. A new approach for constructing distinguishing attacks based on a convolutional neural network is proposed. The approach is based on distinguishing between graphic equivalents of ciphertexts received by the CTR (counter) encryption mode after different number of rounds, including the number of rounds guaranteeing satisfaction of statistical properties. Several schemes are presented for constructing distinguishing attacks, which in some cases make it possible

to detect deviations from randomness in smaller samples than previously known, and with a large number of rounds. The approach allows to create distinguishers without the need for an analytical research of each cipher, which makes it possible to build universal distinguishers for a series of ciphers.

**Keywords:** *block cipher, machine learning, neural network, statistical analysis, distinguishing attack, cryptanalysis.*

## Введение

Итеративные блочные шифры позволяют решать обширный круг задач и являются одним из наиболее значимых и часто используемых классов криптографических алгоритмов. Такие шифры состоят из раундов шифрования — простых итераций, достаточное количество которых обеспечивает требуемый уровень стойкости. Итеративная структура блочных шифров обуславливает выбор подходов к их криптоанализу. Одной из часто применяемых атак является атака-различитель (англ. disinguishing attack), предназначенная для распознавания шифртекстов, полученных после разного числа раундов. При этом важной задачей является поиск максимального числа раундов, при котором возможно построить такую атаку, и уменьшение размера выборки, при котором удаётся обнаружить отклонения от случайности. Эффективные различители представляют интерес как сами по себе, так и в комплексе, когда на их основе создаются алгоритмы вычисления секретных ключей шифрования.

Методы построения атак-различителей можно условно разделить на два класса: аналитические и эмпирические (в основном статистические). Многие аналитические методы базируются на выявлении дифференциальных [1 – 3], линейных [4] или интегральных [5, 6] признаков, описывающих определённые свойства шифртекста после заданного числа раундов ($r$) и называемых $r$-раундовыми характеристиками. На основании этих характеристик разрабатываются алгоритмы вычисления ключа, используемого на $(r + 1)$-м и последующих раундах посредством их полного или частичного перебора. Роль характеристик в этом процессе состоит в том, чтобы отбрасывать неверные пробные ключи (при расшифровании одного $(r+1)$-го раунда с верным ключом $r$-раундовая характеристика выполняется, а при неверном — нет).

Аналитические различители характерны тем, что они позволяют строить атаки на большое количество раундов, когда требуются огромные вычислительные ресурсы, недоступные на практике (например, порядка $2^{128}$ элементарных операций или бит оперативной памяти). Однако поскольку признаки, используемые в таких различителях, обычно тесно связаны с конкретными шифрами, то они не являются универсальными и эффективны только для ограниченного набора шифров. Имеются работы, в которых атаки на блочные шифры и их свойства описываются в общем виде, но, как правило, делается это на довольно высоком уровне абстракции, что не позволяет применить их к серии шифров без дополнительного анализа [7, 8]. При этом, если структура шифра хотя бы частично конкретизирована и позволяет описать класс шифров, то аналитические оценки и атаки можно распространять на такие классы [9 – 11].

Помимо аналитических методов, для оценки стойкости итеративных блочных шифров могут использоваться эмпирические статистические методы, позволяющие осуществить атаку-различитель в ходе эксперимента на выборке, размер которой приемлем для расчётов [12, 13]. Так, в работе [14] предложен и применён для шифра RC6 универсальный подход к вычислению ключа шифрования, где в качестве атаки-различителя выступает критерий хи-квадрат. Для малого числа раундов, когда для распознавания

отклонения от случайности достаточно небольших выборок, атака осуществляется экспериментально, а для большего числа раундов размер выборки экстраполируется аналитически на основе экспериментальных данных. В рамках этого подхода предложены и успешно применены атаки на основе статистических тестов, использующих динамически изменяемые структуры [15 – 17], что позволило повысить их эффективность для ряда шифров [18 – 20].

Достоинством статистических методов является их универсальность, поскольку по одной и той же схеме можно проанализировать серию шифров без учёта особенностей каждого из них. При этом необходимость проведения экспериментальных расчётов накладывает ограничения на размер выборки. Использование свёрточных нейронных сетей имеет потенциал для снижения размера выборки за счёт учёта паттернов, встречающихся в шифртекстах, в то время как статистические тесты принимают решение на основе неких интегральных характеристик, которые хотя и обновляются после каждого выборочного значения, но не рассматривают всю выборку целиком. Технологии машинного обучения уже применяются в криптоанализе, но в основном они связаны с атаками по побочным каналам [21, 22]. Кроме того, многие эффективные атаки в стегоанализе также используют технологии машинного обучения, в том числе ансамблевые классификаторы и метод опорных векторов [24 – 26].

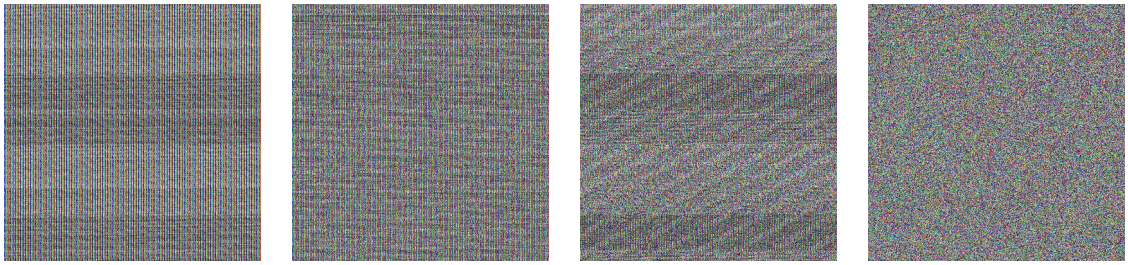
В настоящей работе показано, что свёрточные нейронные сети могут быть использованы для построения универсальных атак-различителей, которые, как демонстрируется экспериментально, в некоторых случаях позволяют выявлять отклонения от случайности на меньших выборках и при большем числе раундов, чем ранее известные статистические тесты.

## 1. Постановка задачи и идея предлагаемого подхода

Задача статистического анализа итеративных блочных шифров состоит в том, чтобы построить атаку-различитель, способную распознать шифртекст после заданного числа раундов, т. е. отличить его от случайной последовательности либо от шифртекста при другом числе раундов. Статистические тесты решают эту задачу посредством вычисления неких интегральных характеристик, которые затем сравниваются с табличными критическими значениями [15 – 17]. При превышении такого значения последовательность признаётся неслучайной с вероятностью $1 - \alpha$, где $\alpha$ — заданный уровень значимости (допустимая вероятность ошибки статистического теста при проверке истинно случайной последовательности). Как правило, с увеличением числа раундов шифрования растёт размер выборки, на котором тест способен отличить шифртекст от случайной последовательности, поэтому этот размер используется для определения числа раундов [12 – 14, 19, 20].

Идея предлагаемого подхода возникла в результате наблюдения, что преобразованный в растровое изображение шифртекст итеративного блочного шифра при разном числе раундов имеет выраженную текстуру (паттерн), которая с увеличением числа раундов изменяется в сторону равномерно шумного (случайного) изображения. Например, на рис. 1 представлены графические эквиваленты шифртекстов итеративного блочного шифра Simon с размером блока 32 бита (Simon-32) после 3, 6, 9 и 30 раундов шифрования.

Рабочая гипотеза нашего исследования заключается в том, что нейронная сеть, решающая задачи классификации изображений, способна различать и шифртексты, полученные при разном числе раундов.



3 раунда 6 раундов 9 раундов 30 раундов

Рис. 1. Различие текстур в графических эквивалентах шифртекстов блочного шифра Simon-32 после разного числа раундов

### 1.1. Инструментарий для проведения экспериментов

Экспериментальное исследование проводилось с помощью нейронной сети Inception-v3 [23], показавшей высокие результаты по распознаванию изображений на конкурсе ImageNet-2014, а предложенная позднее модификация позволила ещё больше увеличить её эффективность.

На основании предварительного анализа для обеспечения приемлемого баланса между скоростью обучения нейронной сети и точностью классификации выбраны следующие параметры: размер изображения составляет 400×400 цветных RGB-пикселей, а размер партии равен 32.

Шифртексты преобразуются в цветные изображения посредством специально разработанной утилиты на языке C++, которая считывает файл с шифртекстом в бинарном виде и записывает каждый байт в качестве значения компонент палитры RGB, формируя 24-битовое растровое BMP-изображение. База шифртекстов для экспериментов формируется с помощью программной библиотеки [27], предназначенной для удобного шифрования сообщений при разном числе раундов.

Размер изображения 400×400 пикселей соответствует выборке размера приблизительно $2^{21{,}9}$ бит: каждый из 160000 пикселей кодируется 24 битами (по 1 байту на каждую компоненту палитры RGB).

Эффективность различителей (способность нейронной сети отличать шифртексты при различном числе раундов друг от друга и от случайных последовательностей) оценивается через долю верных решений нейронной сети на элементах контрольной выборки следующим образом. Пусть $n$ — размер контрольной выборки. Введём следующие случайные величины для $i = 1, \ldots, n$:

$$\eta_i = \begin{cases} 1, \text{если нейронная сеть приняла верное решение} \\ \qquad \text{на } i\text{-м изображении из контрольной выборки}, \\ 0 \quad \text{иначе.} \end{cases}$$

Тогда количество верных решений нейронной сети ($S_n$) и их долю ($\tilde{S}_n$) на всей контрольной выборке можно определить следующим образом:

$$S_n = \sum_{i=1}^{n} \eta_i, \quad \tilde{S}_n = S_n / n.$$

Найдём такое $\tilde{\delta}(n, \alpha)$, что при $\tilde{S}_n \notin [1/2 - \tilde{\delta}(n, \alpha), 1/2 + \tilde{\delta}(n, \alpha)]$ можно сделать вывод о том, что нейронная сеть способна отличать шифртексты и случайные последовательности друг от друга эффективнее простого угадывания.

Пусть

$$S_n^* = \frac{S_n - \mathbb{E}S_n}{\sqrt{\mathbb{D}S_n}}, \tag{1}$$

где $\mathbb{E}S_n$ и $\mathbb{D}S_n$ — математическое ожидание и дисперсия $S_n$ соответственно. Из центральной предельной теоремы следует, что

$$\mathsf{P}\left[S_n^* \in [-\delta, \delta]\right] \approx F_{0,1}(\delta) - F_{0,1}(-\delta),$$

где $F_{0,1}(\cdot)$ — функция стандартного нормального (гауссовского) распределения.

Пусть $Q_{\alpha/2} = F_{0,1}^{-1}(1 - \alpha/2)$ — квантиль стандартного нормального распределения уровня $1 - \alpha/2$, тогда

$$\mathsf{P}\left[S_n^* \in [-Q_{\alpha/2}, Q_{\alpha/2}]\right] \approx 1 - \alpha. \tag{2}$$

Если нейронная сеть не способна отличать графические эквиваленты шифртекстов или случайных последовательностей друг от друга, то все случайные величины $\eta_i$ имеют распределение Бернулли с параметром 1/2, т. е. $\mathsf{P}[\eta_i = 1] = \mathsf{P}[\eta_i = 0] = 1/2$, поскольку результат работы нейронной сети равносилен случайному угадыванию. Следовательно, $\mathbb{E}S_n = n/2$ и $\mathbb{D}S_n = n/4$, а формулу (1) можно преобразовать к виду

$$S_n^* = \frac{S_n - n/2}{\sqrt{n}/2} = \frac{2S_n - n}{\sqrt{n}}. \tag{3}$$

Из формул (2) и (3) получаем

$$\tilde{\delta}(\alpha, n) = \frac{Q_{\alpha/2}}{2\sqrt{n}}. \tag{4}$$

Например, при $\alpha = 0{,}01$ величина $Q_{0{,}01/2}$ равна 2,59, и при таких значениях $\tilde{\delta}(0{,}01, 200) = 0{,}09$ и $\tilde{\delta}(0{,}01, 2000) = 0{,}03$.

## 2. Экспериментальные результаты

Далее представлены результаты экспериментов по различению итеративных блочных шифров при варьируемом числе раундов и случайных последовательностей. Предложено четыре схемы экспериментов, набор которых может быть расширен и другими вариантами.

### 2.1. Базовая схема 1: различение случайной последовательности и шифртекста при сокращённом числе раундов

Задачей первой схемы экспериментов является выявление принципиальной способности нейронной сети отличать случайные последовательности от шифртекста при сокращённом числе раундов. В качестве случайной последовательности взят шифртекст, полученный с помощью 14-раундового шифра AES в режиме счётчика, поскольку многочисленные исследования до настоящего времени не выявили у этого шифра каких-либо уязвимостей, значимых с практической точки зрения. Например, в работе [13] показано, что уже начиная с трёх раундов шифртекст AES обладает удовлетворительными статистическими свойствами. В качестве исследуемого шифра взят шифр Simon-32, поскольку, согласно предварительным экспериментам, статистические свойства его шифртекста достаточно равномерно (без скачков) улучшаются с увеличением числа раундов, что даёт возможность наглядно представить результаты, касающиеся способности нейронной сети находить отклонения от случайности.

Для проведения экспериментов сгенерированы по 1000 шифртекстов алгоритма Simon-32 с различным числом раундов и 1000 случайных последовательностей с помощью 14-раундового AES. Затем для каждого эксперимента нейронная сеть обучалась с целью различения $r$-раундового Simon-32 и случайных последовательностей, $r = 1, \ldots, 10$.

Результаты экспериментов приведены в табл. 1, которая демонстрирует, что при малом числе раундов нейронная сеть приняла 100 % верных решений, а с ростом их числа процент ошибок увеличивается. Тем не менее до 15-го раунда процент ошибок существенно меньше 50 %, т. е. вероятности произвольного угадывания.

Таблица 1

**Оценка способности нейронной сети отличать случайные последовательности от шифртекста при сокращённом числе раундов**

| Число раундов | 3 | 5 | 7 | 9 | 11 | 13 | 15 | 17 | 19 | 21 |
|---|---|---|---|---|---|---|---|---|---|---|
| Доля верных решений | 1,00 | 1,00 | 1,00 | 0,98 | 0,91 | 0,70 | 0,52 | 0,53 | 0,49 | 0,50 |

Поскольку в экспериментах по данной схеме размер контрольной выборки $n = 400$, то при $\alpha = 0{,}01$ по формуле (4) получаем $\tilde{\delta}(0{,}01, 400) \approx 0{,}065$, следовательно, если доля верных решений лежит за пределами интервала [0,43; 0,57], то с вероятностью 0,99 можно считать, что $r$-раундовый шифртекст отличим от случайной последовательности. Таким образом, из результатов табл. 1 можно сделать вывод о том, что при $r \leqslant 15$ блочный шифр Simon не обладает удовлетворительными статистическими свойствами, а при большем числе раундов нейронная сеть способна принимать только решение, равносильное угадыванию.

В работе [12] представлены атаки-различители для шифра Simon-32 (с разными размерами блока) на основе статистического теста «стопка книг» [17]. Максимальное число раундов, при котором выявлены отклонения от случайности у этого шифра, равно 12 на выборке размера $2^{36}$ бит. Этот результат достигнут для 64-битового блока. Для Simon с 32-битовым блоком отклонения от случайности найдены для 9 раундов на выборке размера $2^{27}$. Таким образом, можно сделать вывод, что нейронная сеть способна отличить от случайности большее число раундов и на меньших выборках.

### 2.2. Схема 2: различение соседних раундов

Следующая схема может применяться, например, в отсутствие источника случайных чисел. Она сравнивает графические эквиваленты шифртекстов соседних раундов итеративного блочного шифра. В качестве обучающей выборки на вход подаются зашифрованные алгоритмом Simon-32 последовательности с 3 по 21 раунд (по 400 изображений на каждый). Задача состоит в том, чтобы выяснить, насколько нейронная сеть способна отличать соседние раунды друг от друга.

На рис. 2 представлены результаты экспериментов. Обратим внимание, что линия тренда постепенно сходится к 0,5, т. е. нейронной сети становится сложнее распознавать шифртексты. Найдём число раундов, при котором нейронная сеть способна отличать шифртексты от случайной последовательности в рамках данной схемы. Здесь размер контрольной выборки $n = 200$, значит, $\tilde{\delta}(0{,}01, 200) \approx 0{,}092$, следовательно, если нейронная сеть способна выявлять отклонения, то доля верных решений должна лежать за пределами интервала [0,408; 0,592]. Таким образом, заключаем, что при $r \leqslant 14$ блочный шифр Simon не обладает удовлетворительными статистическими свойствами, а при большем числе раундов нейронная сеть способна только угадывать. Видно, что в рамках данной схемы нейронная сеть работает менее эффективно, чем в рамках

базовой схемы (отличает на 1 раунд меньше), но по-прежнему эффективнее атаки-различителя из работы [12].

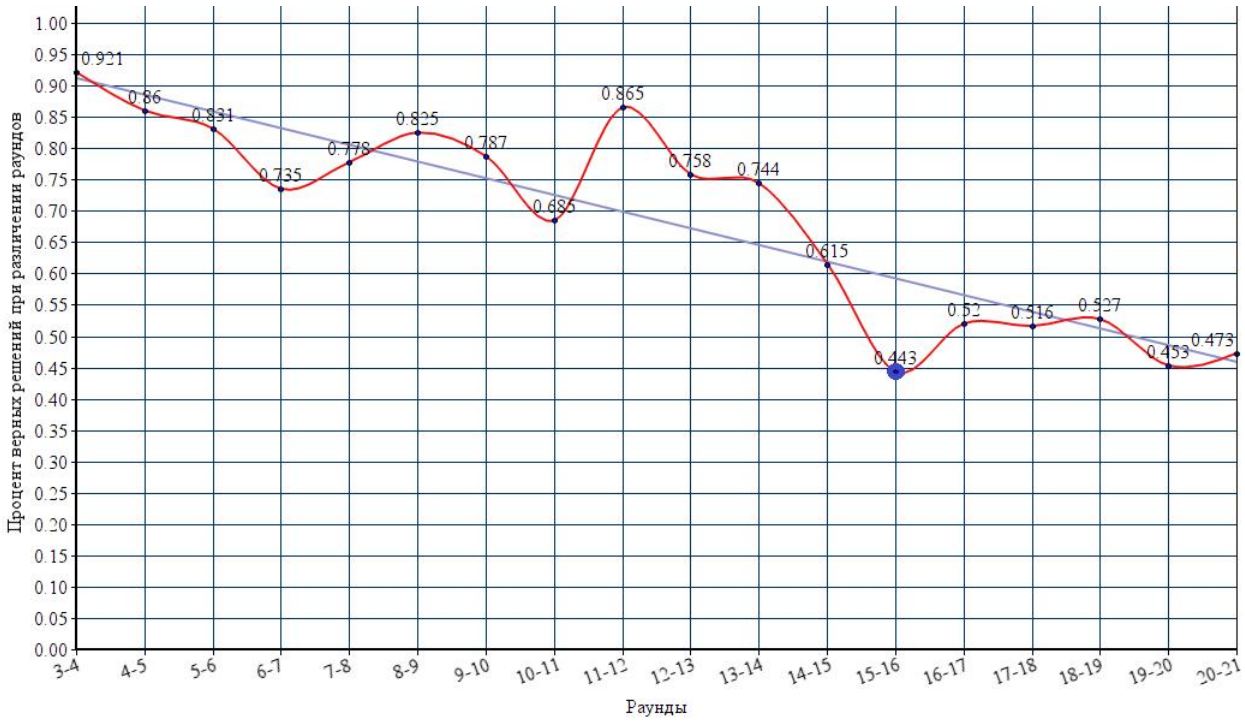



Рис. 2. Оценка способности нейронной сети различать соседние раунды шифра

### 2.3. Схема 3: различение двух шифров

Цель данной схемы — проанализировать, насколько нейронная сеть способна отличать шифры друг от друга. Для экспериментов выбраны блочные шифры Speck-32 и Present, поскольку, согласно предварительному анализу, они демонстрируют похожие статистические свойства при одинаковом числе раундов (наиболее сложный для нейронной сети случай).

Для каждого из анализируемых раундов формируется обучающая выборка размера 800 (включающая по 400 графических эквивалентов шифртекстов каждого из двух шифров при равном числе раундов). Контрольная выборка имеет размер 200 и включает по 100 шифртекстов каждого шифра. Результаты представлены в табл. 2. Как и в схеме 2, здесь $n = 200$, $\tilde{\delta}(0{,}01, 200) = 0{,}092$ и интервал для доли верных решений — [0,408; 0,592].

Таблица 2

**Оценка способности нейронной сети различать шифртексты разных шифров при варьируемом числе раундов**

| Число раундов | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|
| Доля верных решений | 1,00 | 1,00 | 0,99 | 1,00 | 0,91 | 0,66 | 0,59 | 0,49 |

Таким образом, делаем вывод, что нейронная сеть способна различать эти шифры до 10 раундов. Эти результаты напрямую нельзя сравнивать с результатами [12] (решаются разные задачи), где построена атака-различитель для 6 раундов, однако в целом они согласуются. Кроме того, результаты согласуются и с работой [27], где представлен различитель для 8 раундов шифра Present.

### 2.4. Схема 4: различение случаев $r < R_{\min}$ и $r \geqslant R_{\min}$ для нескольких шифров

Данная схема предназначена для того, чтобы выяснить, насколько нейронная сеть способна отличать случайные последовательности от шифртекстов, полученных с помощью различных шифров вперемешку при числе раундов, меньшем, чем найденные в работе [27] значения $R_{\min}$.

Обучающая выборка сформирована из графических эквивалентов 8500 шифртекстов, полученных с помощью 16 шифров (по 500 шифртекстов для каждого) и 500 шифртекстов, полученных с помощью 14-раундового AES, которые считаем случайными последовательностями. Проанализированы шифры LBlock, Present, XTEA, Twine, Speck, Clefia, Hight, Piccolo, Klein, Skipjack, mCrypton, LED, Noekeon, Sea, Mibs, DESXL. Контрольная выборка имеет размер 2000 и составлена из 1000 шифртекстов и 1000 случайных последовательностей.

По результатам экспериментов нейронная сеть приняла верное решение на 0,98 доли выборок. В данном случае $n = 2000$, $\tilde{\delta}(0{,}01, 2000) = 0{,}029$ и интервал — $[0{,}471; 0{,}529]$. Таким образом, доля верных решений лежит за пределами этого интервала и можно сделать вывод о способности нейронной сети находить отклонения от случайности по этой схеме.

## Заключение

Сформулируем основные выводы по итогам экспериментов.

1) Нейронная сеть способна различать шифртексты одного и того же итеративного блочного шифра при полном и сокращённом числе раундов.
2) Нейронная сеть способна различать шифртексты, полученные при шифровании соседними раундами одного блочного шифра.
3) Нейронная сеть способна различать шифртексты, полученные при помощи разных шифров.
4) Нейронная сеть способна отличать шифртексты, полученные при $r < R_{\min}$, от шифртекстов, полученных при $r \geqslant R_{\min}$, в том числе для разных шифров.
5) На примере блочного шифра Simon-32 показано, что для некоторых шифров различители на основе нейронных сетей могут быть эффективнее: требовать меньшую выборку либо работать на большем числе раундов.

В дальнейшем с помощью предложенного подхода можно вычислять $R_{\min}$ и строить алгоритмы вычисления секретного ключа, используя различители на основе нейронных сетей.

## ЛИТЕРАТУРА

1. *Biham E. and Shamir A.* Differential cryptanalysis of DES-like cryptosystems // J. Cryptology. 1991. V. 4. P. 3–72.
2. *Knudsen L.* Truncated and higher order differentials // LNCS. 1994. V. 1008. P. 196–211.
3. *Biham E., Biryukov A., and Shamir A.* Cryptanalysis of Skipjack reduced to 31 rounds using impossible differentials // J. Cryptology. 2005. V. 18. P. 291–311.
4. *Matsui M.* Linear cryptanalysis method for DES cipher // LNCS. 1994. V. 765. P. 386–397.
5. *Knudsen L.* Integral cryptanalysis // LNCS. 2002. V. 2365. P. 112–127.
6. *Biryukov A. and Shamir A.* Structural cryptanalysis of SASAS // J. Cryptology. 2010. V. 23. P. 505–518.

7. *Агибалов Г. П.* Элементы теории дифференциального криптоанализа итеративных блочных шифров с адаптивным раундовым ключом // Прикладная дискретная математика. 2008. № 1(1). С. 34–42.
8. *Денисов О. В.* Критерии марковости алгоритмов блочного шифрования // Прикладная дискретная математика. 2018. № 41. С. 28–37.
9. *Денисов О. В., Былина Р. А.* Матричная формула для распределения выхода блочной схемы шифрования и статистический критерий на ее основе // Прикладная дискретная математика. 2016. № 2(32). С. 33–48.
10. *Токарева Н. Н.* О квадратичных аппроксимациях в блочных шифрах // Проблемы передачи информации. 2008. № 3. С. 105–127.
11. *Агибалов Г. П.* Substitution block ciphers with functional keys // Прикладная дискретная математика. 2017. № 38. С. 57–65.
12. *Сосков А. С., Рябко Б. Я.* Применение атаки различения на легковесные блочные шифры, основанные на ARX-операциях // Вычислительные технологии. 2019. Т. 24. № 3. С. 106–116.
13. *Пестунов А. И.* Статистический анализ современных блочных шифров // Вычислительные технологии. 2007. Т. 12. № 2. С. 122–129.
14. *Knudsen L. and Meier W.* Correlations in RC6 with a reduced number of rounds // LNCS. 2001. V. 1978. P. 94–108.
15. *Рябко Б. Я., Стогниенко В. С., Шокин Ю. И.* Адаптивный критерий хи-квадрат для различения близких гипотез при большом числе классов и его применение к некоторым задачам криптографии // Проблемы передачи информации. 2003. Т. 39. № 2. С. 53–62.
16. *Монарев В. А., Рябко Б. Я.* Экспериментальный анализ генераторов псевдослучайных чисел при помощи нового статистического теста // Журнал вычисл. матем. и матем. физики. 2004. Т. 44. № 5. С. 766–770.
17. *Рябко Б. Я., Пестунов А. И.* «Стопка книг» как новый статистический тест для случайных чисел // Проблемы передачи информации. 2004. Т. 40. № 1. С. 73–78.
18. *Рябко Б. Я., Монарев В. А., Шокин Ю. И.* Новый тип атак на блоковые шифры // Проблемы передачи информации. 2005. Т. 41. № 4. С. 97–107.
19. *Монарев В. А.* Реализация новой статистической атаки на блочный шифр // Вестник СибГУТИ. 2014. № 1. С. 85–90.
20. *Лысяк А. С., Рябко Б. Я., Фионов А. Н.* Анализ эффективности градиентной статистической атаки на блоковые шифры RC6, MARS, CAST-128, IDEA, Blowfish в системах защиты информации // Вестник СибГУТИ. 2013. № 1. С. 85–109.
21. *Lerman L., Bontempi G., and Markowitch O.* A machine learning approach against a masked AES // J. Cryptogr. Eng. 2015. V. 5. P. 123–139.
22. *Hettwer B., Gehrer S., and Guneysu T.* Applications of machine learning techniques in side-channel attacks: a survey // J. Cryptogr. Eng. 2020. V. 10. P. 135–162.
23. *Szegedy C., Vanhoucke V., Ioffe S., et al.* Rethinking the inception architecture for computer vision // Proc. IEEE Conf. CVPR. Las Vegas, NV, USA, June 27–30, 2016. P. 2818–2826.
24. *Монарев В. А., Пестунов А. И.* Эффективное обнаружение стеганографически скрытой информации посредством интегрального классификатора на основе сжатия данных // Прикладная дискретная математика. 2018. № 40. С. 59–71.
25. *Монарев В. А., Пестунов А. И.* Повышение эффективности методов стегоанализа при помощи предварительной фильтрации контейнеров // Прикладная дискретная математика. 2016. №. 2(32). С. 87–99.
26. *Kodovsky J., Fridrich J., and Holub V.* Ensemble classifiers for steganalysis of digital media // IEEE Trans. Inform. Forensics and Security. 2010. V. 7. No. 2. P. 434–444.

27. *Пестунов А. И., Перов А. А.* Программная библиотека для статистического анализа итеративных блочных шифров // Информационное противодействие угрозам терроризма. 2015. № 24. С. 197–202.

## REFERENCES

1. *Biham E. and Shamir A.* Differential cryptanalysis of DES-like cryptosystems. J. Cryptology, 1991, vol. 4, pp. 3–72.
2. *Knudsen L.* Truncated and higher order differentials. LNCS, 1994, vol. 1008, pp. 196–211.
3. *Biham E., Biryukov A., and Shamir A.* Cryptanalysis of Skipjack reduced to 31 rounds using impossible differentials. J. Cryptology, 2005, vol. 18, pp. 291–311.
4. *Matsui M.* Linear cryptanalysis method for DES cipher. LNCS, 1994, vol. 765, pp. 386–397.
5. *Knudsen L.* Integral cryptanalysis. LNCS, 2002, vol. 2365, pp. 112–127.
6. *Biryukov A. and Shamir A.* Structural cryptanalysis of SASAS. J. Cryptology, 2010, vol. 23, pp. 505–518.
7. *Agibalov G. P.* Elementy teorii differencial'nogo kriptoanaliza iterativnyh blochnyh shifrov s adaptivnym raundovym klyuchom [Some theoretical aspects of differential cryptanalysis of the iterated block ciphers with additive round key]. Prikladnaya Diskretnaya Matematika, 2008, no. 1(1), pp. 34–42. (in Russian)
8. *Denisov O. V.* Kriterii markovosti algoritmov blochnogo shifrovaniya [Markov criteria for block cipher algorithms]. Prikladnaya Diskretnaya Matematika, 2018, no. 41, pp. 28–37. (in Russian)
9. *Denisov O. V. and Bylina R. A.* Matrichnaya formula dlya raspredeleniya vyhoda blochnoj skhemy shifrovaniya i statisticheskij kriterij na ee osnove [Matrix formula for the spectrum of output distribution of block cipher scheme and statistical criterion based on this formula]. Prikladnaya Diskretnaya Matematika, 2016, no. 2(32), pp. 33–48. (in Russian)
10. *Tokareva N. N.* O kvadratichnyh approksimaciyah v blochnyh shifrah [About quadratic approximations in block ciphers]. Problemy Peredachi Informacii, 2008, vol. 3, pp. 105–127. (in Russian)
11. *Agibalov G. P.* Substitution block ciphers with functional keys. Prikladnaya Diskretnaya Matematika, 2017, no. 38, pp. 57–65.
12. *Soskov A. S. and Ryabko B. Ya.* Primenenie ataki razlicheniya na legkovesnye blochnye shifry, osnovannye na ARX-operaciyah [Applying distinction attack on lightweight block ciphers based on ARX operations] Vychislitel'nye Tekhnologii, 2019, vol. 3, pp. 106–116. (in Russian)
13. *Pestunov A. I.* Statisticheskij analiz sovremennyh blochnyh shifrov [Statistical analysis of modern block ciphers]. Vychislitel'nye Tekhnologii, 2007, vol. 12, no. 2, pp. 122–129. (in Russian)
14. *Knudsen L. and Meier W.* Correlations in RC6 with a reduced number of rounds. LNCS, 2001, vol. 1978, pp. 94–108.
15. *Ryabko B. Ya., Stognienko V. S., and SHokin Yu. I.* Adaptivnyj kriterij hi-kvadrat dlya razlicheniya blizkih gipotez pri bol'shom chisle klassov i ego primenenie k nekotorym zadacham kriptografii [Adaptive Chi-square test for distinguishing close hypotheses with a large number of classes and its application to some cryptography problems]. Problemy Peredachi Informacii, 2003, vol. 39, no. 2, pp. 53–62. (in Russian)
16. *Monarev V. A. and Ryabko B. Ya.* Eksperimental'nyj analiz generatorov psevdosluchajnyh chisel pri pomoshchi novogo statisticheskogo testa [Experimental analysis of pseudo-random number generators using a new statistical test]. Zhurnal Vychislitel'noj Matematiki i Matematicheskoj Fiziki, 2004, vol. 44, no. 5, pp. 766–770. (in Russian)

17. *Ryabko B. Ya. and Pestunov A. I.* “Stopka knig” kak novyj statisticheskij test dlya sluchajnyh chisel [Book Stack as a new statistical test for random numbers]. Problemy Peredachi Informacii, 2004, vol. 40, no. 1, pp. 73–78. (in Russian)
18. *Ryabko B. Ya., Monarev V. A., and Shokin Yu. I.* Novyj tip atak na blokovye shifry [A new type of attack on block ciphers]. Problemy Peredachi Informacii, 2005, vol. 41, no. 4, pp. 97–107. (in Russian)
19. *Monarev V. A.* Realizaciya novoj statisticheskoj ataki na blochnyj shifr [Implementation of a new statistical attack on a block cipher]. Vestnik SibGUTI, 2014, vol. 1, pp. 85–90. (in Russian)
20. *Lysyak A. S., Ryabko B. Ya., and Fionov A. N.* Analiz effektivnosti gradientnoj statisticheskoj ataki na blokovye shifry RC6, MARS, CAST-128, IDEA, Blowfish v sistemah zashchity informacii [Analysis of the effectiveness of gradient statistical attacks on block ciphers RC6, MARS, CAST-128, IDEA, Blowfish in information security systems]. Vestnik SibGUTI, 2013, vol. 1, pp. 85–109. (in Russian)
21. *Lerman L., Bontempi G., and Markowitch O.* A machine learning approach against a masked AES. J. Cryptogr. Eng., 2015, vol. 5, pp. 123–139.
22. *Hettwer B., Gehrer S., and Guneysu T.* Applications of machine learning techniques in side-channel attacks: a survey. J. Cryptogr. Eng., 2020, vol. 10, pp. 135–162.
23. *Szegedy C., Vanhoucke V., Ioffe S., et al.* Rethinking the inception architecture for computer vision // Proc. IEEE Conf. CVPR, Las Vegas, NV, USA, June 27–30, 2016, pp. 2818–2826.
24. *Monarev V. A. and Pestunov A. I.* Effektivnoe obnaruzhenie steganograficheski skrytoj informacii posredstvom integral’nogo klassifikatora na osnove szhatiya dannyh [Efficient steganography detection by means of compression-based integral classifier]. Prikladnaya Diskretnaya Matematika, 2018, no. 40, pp. 59–71. (in Russian)
25. *Monarev V. A. and Pestunov A. I.* Povyshenie effektivnosti metodov stegoanaliza pri pomoshchi predvaritel’noj fil’tracii kontejnerov [Enhancing steganalysis accuracy via tentative filtering of stego-containers]. Prikladnaya Diskretnaya Matematika, 2016, no. 2(32), pp. 87–99. (in Russian)
26. *Kodovsky J., Fridrich J., and Holub V.* Ensemble classifiers for steganalysis of digital media. IEEE Trans. Information Forensics and Security, 2010, vol. 7, no. 2, pp. 434–444.
27. *Pestunov A. I. and Perov A. A.* Programmnaya biblioteka dlya statisticheskogo analiza iterativnyh blochnyh shifrov [Software library for statistical analysis of iterative block ciphers]. Informacionnoe Protivodejstvie Ugrozam Terrorizma, 2015, vol. 24, pp. 197–202. (in Russian)

УДК 004.056.5

# ОЦЕНКА ВЕРОЯТНОСТИ ВЫИГРЫША ПРИ ПРОВЕДЕНИИ МАЙНИНГА НЕБОЛЬШОЙ ГРУППОЙ УЧАСТНИКОВ

А. В. Черемушкин

*НИИ «Квант», г. Москва, Россия*

В работах Ittay Eyal и Emin Gün Sirer показано, что протокол майнинга, реализованный в биткоине, является уязвимым к атаке со стороны группы участников, составляющей относительно небольшую часть от общего числа майнеров, позволяющей ей получить вознаграждение, превышающее размер доли имеющихся у них вычислительных ресурсов, и описана стратегия проведения т. н. корыстного майнинга. В данной работе описана уточнённая вероятностно-автоматная марковская модель корыстного майнинга, основанная на предположении о независимости обеих групп участников. Пусть доля вычислительных ресурсов у корыстной группы пропорциональна $p$, $0 < p < 1/2$, а у второй группы — $(1 - p)$. Рассматривается также ситуация, когда в случае разветвления цепочки блоков во второй группе часть участников, пропорциональная $\gamma(1-p)$, будет строить продолжение для цепочки, сформированной первой группой, а остальные (относительная доля $(1 - \gamma)(1 - p)$) — для второй цепочки. Основной результат состоит в обосновании уточнённого интервала $0 < p \leqslant 0{,}429$, соответствующего значениям параметра $p$, при котором корыстная группа получает относительное вознаграждение, превышающее вознаграждение при честном майнинге. Левая граница соответствует значению $\gamma = 1$, а правая — 0. Аналогично, при $0{,}358 \leqslant p \leqslant 0{,}454$ и подходящих значениях $\gamma$ корыстная группа получает относительное вознаграждение, превышающее вознаграждение остальных участников.

**Ключевые слова:** *блокчейн, майнинг, марковская модель, вероятностный автомат.*

DOI 10.17223/20710410/49/5

# SELFISH MINING STRATEGY ELABORATION

A. V. Cheremushkin

*"Research Institute Kvant", Moscow, Russia*

**E-mail:** avc238@mail.ru

As it was shown by Ittay Eyal and Emin Gün Sirer, the Bitcoin mining protocol is not incentive-compatible, because there exists an attack in which colluding miners obtain a revenue larger than their fair share. We describe an elaboration of Selfish-Mine Strategy and present an extended model of selfish mining based on independency hypothesis: both groups are made their work independently from each other. We describe a new state machine modelling selfish pool strategy. Let the selfish pool has mining power of $p$, $0 < p < 1/2$, and the others of $(1 - p)$. We also consider the situation in which the others mine a block on the previously private branch (frequency $\gamma(1-p)$), and the others mine a block on the public branch (frequency $(1-\gamma)(1-p)$). Main result is an elaboration of an interval in which selfish miners will earn more than their relative mining power: 1) for a given $p$, a pool of size $p$ obtains a revenue larger

than than its relative size for $p$ in the following range: $0 < p \leqslant 0.429$ (the left bound corresponds to $\gamma = 1$, and the right one — to $\gamma = 0$); 2) for a given $p$, a pool of size $p$ obtains a revenue larger than a revenue of other group in the following range: $0.358 \leqslant p \leqslant 0.449$.

**Keywords:** *blockchain, mining, Markov model, state machine.*

## Введение

В работе [1] анонсирована, а позднее в [2, 3] опубликована стратегия поведения выделенной (корыстной) группы участников майнинга, у которой суммарная вычислительная мощность принадлежащих им ресурсов не превосходит половины от общей вычислительной мощности, используемой для майнинга. Стратегия, названная авторами *стратегией корыстного майнинга*, позволяет получать данной группе вознаграждение, превышающее размер доли имеющихся у них вычислительных ресурсов. В её основе лежит утаивание части успешно подобранных блоков, с помощью которых можно уменьшить выигрыш остальной группы участников. Для обоснования предложенной стратегии авторы описали модель поведения всей системы, в рамках которой можно показать, что стратегия даёт выигрыш даже в случае, когда доля вычислительных ресурсов этой группы может составлять всего $1/3$ от общей суммарной вычислительной мощности всех участников майнинга.

В настоящей работе приводится уточнение данной модели, основанное на модели независимого поведения обеих групп.

## 1. Модель протокола формирования блокчейна

Напомним основные моменты способа формирования цепочки блоков (блокчейна) на основе поиска решения трудоемкой задачи. В криптовалюте Биткоин в качестве такой задачи выбрана задача подбора случайного вектора, который на входе хеш-функции вместе с хеш-свёрткой данных о транзакциях, входящих в данный блок, и свёрткой предыдущего блока даёт значение, имеющее в старших разрядах в двоичной записи $n$ нулей, где $n$ — большое число, определяющее сложность решения задачи. Средняя трудоёмкость $M$ решения задачи подбора такого случайного вектора составляет $2^n$ операций хеширования, поскольку при случайном выборе аргумента вероятность появления на выходе хеш-функции такого значения составляет $2^{-n}$. Поиск такого случайного вектора называется майнингом, которым занимается большое число майнеров, объединяющихся в группы (пулы) для увеличения вероятности успешного подбора. Разные пулы могут формировать различные наборы транзакций для очередного блока. Это является одной из причин появления разветвлений в блокчейне, когда появляется несколько продолжений для уже имеющейся цепочки блоков.

Параметры майнинга в биткоине обычно рассчитываются таким образом, чтобы один блок генерировался в среднем каждые 10 м. При фиксированной суммарной вычислительной мощности этот временной интервал $t$ рассчитывается как математическое ожидание геометрического распределения $t = 1/P$. Здесь вероятность успешного подбора за секунду определяется равенством $P = V/M$, где $V$ — производительность, т. е. число операций хеширования в секунду. Если суммарная вычислительная мощность $V$ возрастает, то для сохранения этого интервала увеличивается трудоёмкость $M$ подбора соответствующего значения хеш-свертки, т. е. параметр $n$.

В результате можно считать, что вероятность успешного формирования блока в этом временном интервале равна единице.

В случае, когда нескольким пулам удается успешно сфомировать продолжение для последнего блока, публикуется разветвление. Далее осуществляется поиск продолжения для каждого из вариантов и в конечном счёте оставляется то, которое имеет наибольшую длину. В этом случае вознаграждение получают те пулы, которые участвовали в формировании наиболее длинной цепочки. Чтобы ограничить время принятия, решения обычно ограничивают длину цепочек в разветвлении (например, шестью блоками). Будем рассматривать наиболее простую модель, когда длины цепочек в разветвлении не ограничены, а решение об оставлении принимается при простом превышении длин на 1; при этом пренебрегаем временем на пересылку данных и временем принятия решения.

## 2. Описание модели из [1]

Предложенный в [1] оригинальный метод основан на использовании теоретико-автоматной марковской модели поведения системы, представляющей собой вероятностный автомат со счётным множеством состояний $s = 0, 1, \ldots$, в котором к состоянию $s = i$ относятся все ситуации, в которых корыстная группа имеет преимущество в числе правильно сформированных блоков, сохраняемых в тайне и образующих цепочку из $i$ блоков. При этом рассматриваются только те состояния, когда $i \geqslant 0$, поскольку в случае, когда группа остальных участников получает преимущество в числе правильно сформированных блоков, корыстная группа прекращает формирование своей цепочки, присоединяется к остальным участникам и начинает подбирать продолжение для цепочки, сформированной остальными участниками (тем самым осуществляется переход в состояние $s = 0$). В автомате введено ещё одно дополнительное состояние $0'$, соответствующее разветвлению на две цепочки одинаковой длины равной 1.

Пусть $\alpha$, $0 < \alpha < 1/2$, — доля в суммарной производительности вычислительных ресурсов $V$, которой обладает группа корыстных участников майнинга. Тогда вероятность успешного формирования блока корыстной группой в течение временного интервала $t$ равна $p = \alpha$, а для группы остальных участников соответственно $q = 1 - \alpha$. Для вычисления вероятностей нахождения автомата в каждом из состояний авторы [1] определяют вероятности перехода, исходя из предположения, что каждый раз успеха достигает либо одна, либо другая группа участников.

Идея состоит в том, что в случае успешного подбора очередного блока корыстная группа не обнародует результат, а держит его в тайне от остальных до тех пор, пока остальные участники сами не подберут очередной блок. В этом случае она поступает одним из следующих вариантов:

— если $s = 1$, то они обнародуют свой блок, создавая разветвление длины 1 и откладывая вопрос о том, какая из групп получит вознаграждение;
— если $s = 2$, то корыстная группа раскрывает оба своих блока, тем самым получая вознаграждение за два блока и лишая вознаграждения остальных участников;
— если $s \geqslant 3$, то они обнародуют блок, стоящий в начале своей сохраняемой в тайне цепочки, создавая разветвление из двух цепочек длины 1 либо увеличивая на 1 длину цепочки в существующем разветвлении, тем самым лишая группу остальных участников выигрыша.

Авторы [1] рассмотрели также случай, когда при наличии разветвления среди остальных участников найдется подгруппа, составляющая (по мощности) долю, равную $\gamma$, $0 \leqslant \gamma \leqslant 1$, которая будет пытаться продолжить ветку, созданную выделенной группой, тем самым повышая вероятность получения вознаграждения корыстной группой за блоки, подобранные ею ранее. Граф переходов вероятностного автомата,

моделирующего поведение участников, приведён на рис. 1. В алгоритме 1 [1] приняты следующие обозначения: $L_{\text{priv}}$ — длина приватной цепочки, сохраняемой в тайне корыстной группой; $L_{\text{publ}}$ — длина общедоступной цепочки в разветвлении.

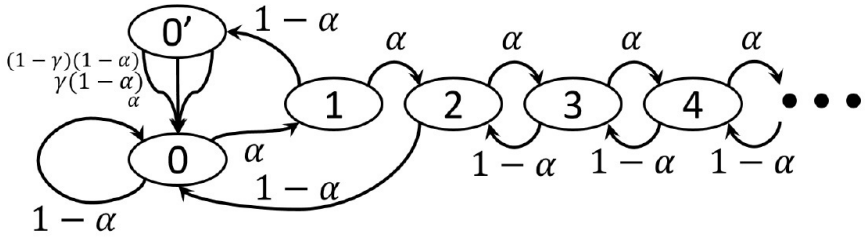



Рис. 1. Автоматная модель из [1]

**Алгоритм 1.** [1]

**Инициализация**
Общедоступная цепочка ← общеизвестные блоки.
Приватная цепочка ← общеизвестные блоки.
$L_{\text{priv}} := 0$.
Майнинг нового блока приватной цепочки.
**Корыстная группа нашла блок**
$\Delta_{\text{prev}} := L_{\text{priv}} - L_{\text{publ}}$.
Добавление нового блока к приватной цепочке.
$L_{\text{priv}} := L_{\text{priv}} + 1$.
**Если** $\Delta_{\text{prev}} = 0$ и $L_{\text{priv}} = 2$, **то**
    публикация всей приватной цепочки. *// Корыстная группа выигрывает благодаря большей длине на 1*
    $L_{\text{priv}} := 0$.
Майнинг нового блока приватной цепочки.
**Другие нашли блок**
$\Delta_{\text{prev}} := L_{\text{priv}} - L_{\text{publ}}$.
Добавление нового блока к общедоступной цепочке.
**Если** $\Delta_{\text{prev}} = 0$, **то**
    приватная цепочка ← общеизвестная цепочка. *// Вторая группа выиграла*
    $L_{\text{priv}} := 0$.
**иначе**
    **Если** $\Delta_{\text{prev}} = 1$, **то**
        публикация последнего блока приватной цепочки. *// Цепочки имеют одинаковую длину, разветвление длины 1*
    **иначе**
        **Если** $\Delta_{\text{prev}} = 2$, **то**
            публикация всей приватной цепочки. *// Корыстная группа выигрывает благодаря большей длине цепочки на 1*
            $L_{\text{priv}} := 0$,
            **иначе** $(\Delta_{\text{prev}} > 2)$
            Публикация первого неопубликованного блока приватной цепочки.
Майнинг нового блока приватной цепочки.

Рассмотренная в [1] модель позволяет успешно рассчитать вероятность получения вознаграждения, превышающего размер вознаграждения при честном майнинге: корыстная группа получает преимущество при выполнении неравенства

$$\frac{1-\gamma}{3-2\gamma} < \alpha < \frac{1}{2}.$$

## 3. Особенности предложенной в [1] модели

В [1] строится вероятностно-автоматная модель для случая, когда все майнеры разделены на две параллельно работающие группы, причём каждый раз успеха достигает либо одна, либо другая группа. Поэтому можно считать, что при переходе автоматной модели из одного состояния в другое производится бросание одной монеты с вероятностями выпадания сторон $p = \alpha$ и $q = 1 - \alpha$, а функционирование вероятностного автомата определяется последовательностью независимых испытаний с двумя исходами. Поскольку вероятности $p$ и $q$ определены для временного интервала $t$, для которого $p + q = 1$, то в рамках данной модели необходимо, чтобы выполнялось условие: после того как одна из групп достигает успеха, обе группы должны останавливать свои вычислительные устройства и начинать майнинг заново.

Обратим внимание на две особенности рассматриваемого подхода.

Во-первых, если бы обе группы придерживались стратегии честного майнинга и сразу объявляли об успешно сформированном блоке, то они бы одновременно начинали новый поиск очередного блока. Поэтому такую ситуацию можно было бы моделировать пуассоновским процессом с непрерывным временем с интенсивностью $P$, а моменты перехода автомата из одного состояния в другое совпадали бы с моментами останова компьютеров при успехе одной из групп. При этом выигрыш каждой из групп был бы пропорционален имеющимся вычислительным мощностям.

При использовании стратегии корыстного майнинга малая группа начинает утаивать результаты и синхронизацию времени по моментам подбора блоков нельзя произвести. Дело в том, что хотя бóльшая группа сразу объявляет о построенном блоке, малая группа может придерживать найденные блоки до тех пор, пока бóльшая группа не достигнет успеха, а затем раскрывать их, чтобы нейтрализовать выигрыш большой группы. Поэтому возможна ситуация, когда бóльшая группа ещё не смогла подобрать нужного хеш-значения и продолжает майнинг, а первая подобрала и утаила. С другой стороны, если бóльшая группа построила блок, но у меньшей группы есть несколько блоков в запасе, то она раскрывает первый блок своей цепочки для нейтрализации выигрыша другой стороны, но сама продолжает майнинг для поиска продолжения для последнего блока в своей цепочке. Поэтому моменты начала майнинга нового блока у разных групп разные.

Но граф переходов на рис. 1 строится исходя из условия, что при переходе в каждое состояние обе группы начинают подбор соответствующего блока заново, останавливая предыдущий поиск. Это делает данную модель моделью с дискретным временем. Поэтому в исходной работе, фактически, применяется более простая модель с дискретным временем, в которой синхронизация производится по фиксированным временным интервалам, в каждом из которых обязательно должен быть найден блок, причём числа $p$ и $q$ обозначают вероятности успешного подбора блока для корыстной группы и группы остальных участников в течение этого временного интервала.

Дискретная модель является, безусловно, огрублением реальной ситуации, когда время до следующего успешного формирования блока является случайной величиной.

Однако с её помощью можно просто оценивать вероятности успеха. Модель с дискретным временем часто используется для предварительной оценки ситуации. Главная идея авторов состоит в построении компактной вероятностно-автоматной модели, состояниями которой являются такие ситуации, когда первая группа имеет преимущество в некотором фиксированном числе сформированных блоков, сохраняемых в тайне от большей группы.

Вторая особенность заключается в следующем. На самом деле, две группы майнеров используют разные стратегии и решают разные задачи, действуя независимо. Это объясняется следующими соображениями. Поскольку группы используют разные стратегии и при этом одна из групп может скрывать результаты, то происходит параллельное решение двух задач двумя группами, обладающими разными вычислительными мощностями. Так как мощность у каждой из групп меньше общей суммарной, то теперь возможна ситуация, когда обе группы не успеют в течение заданного временного интервала подобрать нужного хеш-значения, что соответствует реальной ситуации, когда очередные блоки подбираются не в текущем, а в следующих временных интервалах. Также возможна ситуация, когда обе группы найдут искомые значения, одно из которых может быть скрыто корыстной группой. Если к концу временного интервала объявляется о нахождении двух разных продолжений цепочки блоков, причём одно из них оказывается длиннее, то для включения в блокчейн выбирается более длинная ветвь в разветвлении.

Поэтому более естественно эту ситуацию моделировать последовательностью независимых одинаково распределённых пар случайных величин $(\xi_1, \xi_2)$ с вероятностями успеха $\mathsf{P}[\xi_1 = 1] = p$ (для первой группы) и $\mathsf{P}[\xi_2 = 1] = q$ (для второй группы).

Учтём это замечание и исследуем, как изменится модель для случая двух активных участников.

## 4. Уточнённый подход

Пусть, как и раньше, $p$ — вероятность успеха в подборе нужного значения хеш-функции для корыстной группы, $0 < p < 1/2$, и $q = 1 - p$ — для группы остальных участников. Предполагаем, что обе группы являются однородными и моделируются двумя вычислителями с суммарными мощностями, пропорциональными данным вероятностям. Будем полагать, что обе группы действуют независимо, при этом попрежнему используется модель с дискретным временем. Рассмотрим автономный вероятностный автомат без выхода, множество состояний которого состоит из трёх групп. Первую группу составляют состояния $s_i$, $i = 0, 1, 2, \ldots$, в которых у корыстной группы участников имеется преимущество в числе подобранных хеш-значений для блоков, равное номеру состояния. Отрицательные значения, как и в предыдущей модели, не рассматриваются, так как они соответствуют нулевому состоянию. Вторую группу составляют состояния $s_{i,0}$, $i \geqslant 2$, в которых блокчейн допускает разветвление с двумя продолжениями и длина цепочки, сформированной корыстной группой, содержит на $i$ блоков больше, чем цепочка, сформированная группой остальных участников майнинга. Случай $i = 1$ также не рассматривается, так как в этом случае первая группа раскрывает свою цепочку и система переходит в нулевое состояние. Третью группу образуют состояния $s_{i,i}$ при $i \geqslant 1$, которые соответствуют случаю разветвлений с двумя одинаковыми длинами продолжений исходной цепочки.

Стратегию поведения выделенной группы участников определим аналогично описанному в работе [1] алгоритму (алгоритм 2).

**Алгоритм 2.** Новый алгоритм

**Инициализация**
Общедоступная цепочка ← общеизвестные блоки.
Приватная цепочка ← общеизвестные блоки.
$L_{\text{priv}} := 0$.
Майнинг с головного блока приватной цепочки.
**Корыстная группа нашла блок**
$\Delta_{\text{prev}} := L_{\text{priv}} - L_{\text{publ}}$.
Добавление нового блока к приватной цепочке.
$L_{\text{priv}} := L_{\text{priv}} + 1$.
**Если** $\Delta_{\text{prev}} = 0$ и $L_{\text{priv}} \geqslant 2$, **то**
  публикация всей приватной цепочки, *// Корыстная группа выигрывает благодаря большей длине цепочки на 1*
  $L_{\text{priv}} := 0$.
Майнинг нового блока приватной цепочки.
**Другие нашли блок**
$\Delta_{\text{prev}} := L_{\text{priv}} - L_{\text{publ}}$.
Добавление нового блока к общедоступной цепочке.
**Если** $\Delta_{\text{prev}} = 0$, **то**
  приватная цепочка ← общеизвестная цепочка, *// Они выиграли*
  $L_{\text{priv}} := 0$;
**иначе**
  **Если** $\Delta_{\text{prev}} = 1$, **то**
    публикация последнего блока приватной цепочки, *// Цепочки имеют одинаковую длину, разветвление длины 1*
  **иначе**
    **Если** $\Delta_{\text{prev}} = 2$, **то**
      публикация всей приватной цепочки, *// Корыстная группа выигрывает благодаря большей длине цепочки на 1*
      $L_{\text{priv}} := 0$;
      **иначе** $(\Delta_{\text{prev}} > 2)$
      публикация первого неопубликованного блока приватной цепочки.
**Корыстная группа нашла блок и другие нашли блок**
$\Delta_{\text{prev}} := L_{\text{priv}} - L_{\text{publ}}$.
Добавление нового блока к приватной цепочке.
Добавление нового блока к общедоступной цепочке.
**Если** $\Delta_{\text{prev}} = 0$, **то**
  публикация новых блоков для обеих цепочек,
  $L_{\text{priv}} := L_{\text{priv}} + 1$; $L_{\text{publ}} := L_{\text{publ}} + 1$, *// Цепочки имеют одинаковую длину*
**иначе**
  **Если** $\Delta_{\text{prev}} = 1$ $(L_{\text{priv}} = 2)$, **то**
    публикация всей приватной цепочки, *// Корыстная группа выигрывает благодаря большей длине цепочки на 1*
    $L_{\text{priv}} := 0$;
    **иначе** $(\Delta_{\text{prev}} > 1)$
    публикация последнего блока приватной цепочки. *// Корыстная группа получает отложенный выигрыш за один блок*
Майнинг нового блока приватной цепочки.

Рассмотрим граф переходов вероятностного автомата, моделирующего поведение двух групп участников (рис. 2). В нём вершины помечены индексами соответствующих состояний, а переходы — парами $ab$, соответствующими значениям случайных величин $\xi_1 = a$ и $\xi_2 = b$ $(a, b \in \{0, 1\})$. Метки переходов приведены в табл. 1.

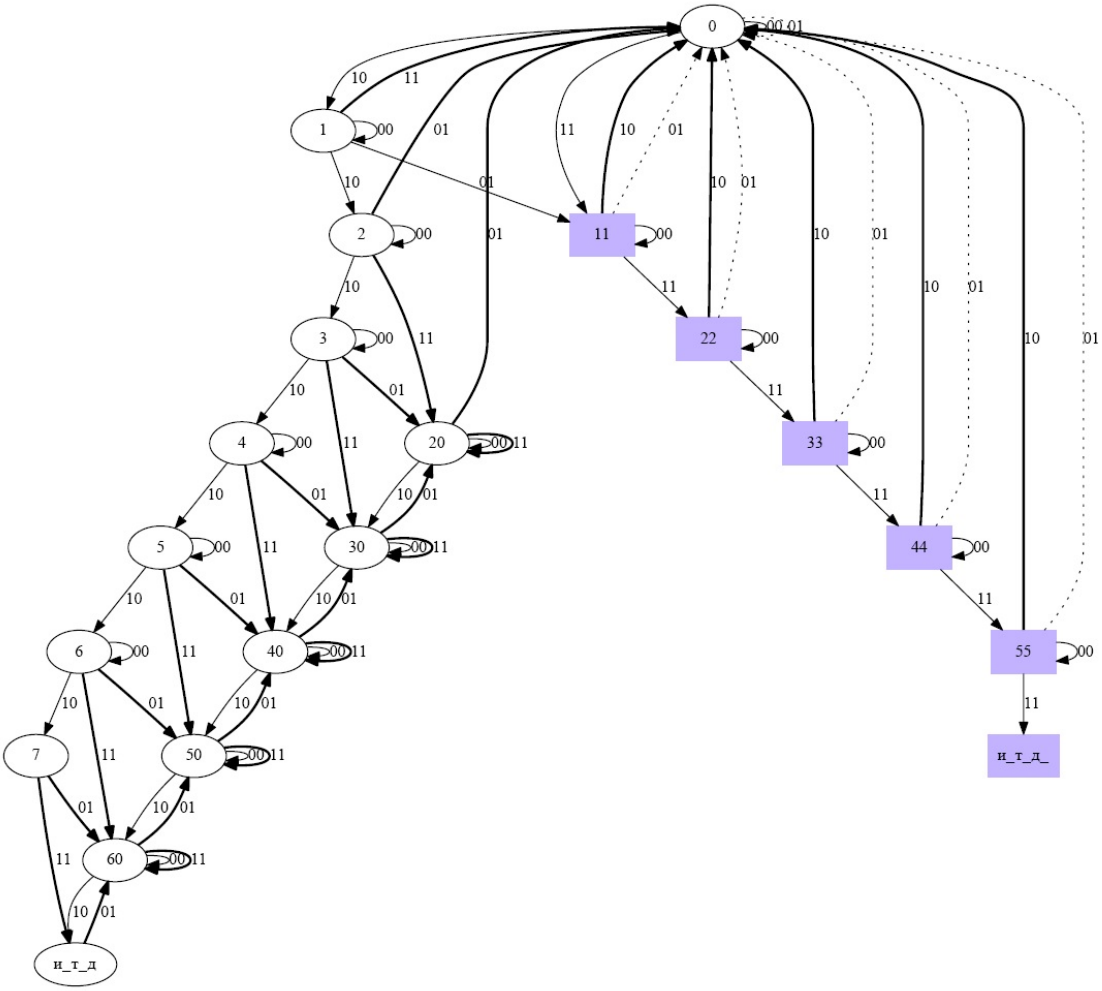



Рис. 2. Граф переходов состояний для уточнённой модели

Таблица 1

**Метки переходов в уточнённой модели**

| Метка | Вероятность перехода | Событие |
|---|---|---|
| 00 | $qp$ | Ни одна группа не нашла продолжения |
| 01 | $q^2$ | Вторая группа нашла продолжение второй цепочки |
| 10 | $p^2$ | Корыстная группа нашла продолжение своей цепочки |
| 11 | $pq$ | Обе группы нашли продолжение для своих цепочек |

Основное отличие в графах переходов на рис. 1 и 2, помимо увеличения до четырёх числа исходящих рёбер из каждой вершины, заключается в том, что состояние $0'$ в первом графе, соответствующее разветвлению на две цепочки одинаковой длины, заменено на цепочку состояний $s_{11}, s_{22}, \ldots$, соответствующих разветвлениям с двумя

одинаковыми цепочками длин $1, 2, \ldots$ соответственно. В первой модели такое разветвление может иметь только длину равную 1, а ситуаций с большими длинами цепочек не возникает, поскольку каждый раз выигрывает либо одна, либо другая сторона, но не обе вместе. Кроме того, добавлены состояния $s_{i0}$ при $i \geqslant 2$, соответствующие разветвлению на две цепочки разных длин, в котором корыстная группа имеет преимущество, равное $i$ блокам. На рис. 2 для изображения рёбер используются линии трёх типов: жирной линией нарисованы рёбра, соответствующие событиям, в которых корыстная группа гарантирует для себя вознаграждение; пунктиром — в которых вознаграждение получает группа остальных участников, а тонкие линии указывают, что ни одна из групп ничего не получает.

На самом деле участники обеих групп получают вознаграждение только при переходе в нулевое состояние, когда соответствующие цепочки окончательно включаются в блокчейн. Поэтому здесь, как и в работе [1], производится приписывание соответствующего вознаграждения переходам графа в тех случаях, когда участники гарантировали себе определённую часть вознаграждения, независимо от дальнейшей траектории поведения.

Поведение автомата моделируется цепью Маркова, для которой переходные вероятности задаются следующим образом. Из каждого состояния возможен переход в четыре состояния в зависимости от того, достигают или нет успеха указанные группы участников. Для удобства варианты переходов для всех состояний собраны в одну таблицу (табл. 2).

Таблица 2

**Переходы графа для уточнённой модели**

| Исходное состояние | Метка ребра | Вероятность перехода | Следующее состояние | Выигрыш обеих групп |
|---|---|---|---|---|
| $s_i$ $(i \geqslant 0)$ | 00 | $qp$ | $s_i$ | $(0,0)$ |
| $s_0$ | 01 | $q^2$ | $s_0$ | $(0,1)$ |
| $s_1$ | 01 | $q^2$ | $s_{1,1}$ | $(0,0)$ |
| $s_2$ | 01 | $q^2$ | $s_0$ | $(2,0)$ |
| $s_i$ $(i \geqslant 3)$ | 01 | $q^2$ | $s_{i-1,0}$ | $(1,0)$ |
| $s_i$ $(i \geqslant 0)$ | 10 | $p^2$ | $s_{i+1}$ | $(0,0)$ |
| $s_0$ | 11 | $pq$ | $s_{1,1}$ | $(0,0)$ |
| $s_1$ | 11 | $pq$ | $s_0$ | $(2,0)$ |
| $s_i$ $(i \geqslant 2)$ | 11 | $pq$ | $s_{i,0}$ | $(1,0)$ |
| $s_{i,i}$ $(i \geqslant 1)$ | 00 | $qp$ | $s_{i,i}$ | $(0,0)$ |
| $s_{i,i}$ $(i \geqslant 1)$ | 01 | $q^2$ | $s_0$ | $(0,i+1)$ |
| $s_{i,i}$ $(i \geqslant 1)$ | 10 | $p^2$ | $s_0$ | $(i+1,0)$ |
| $s_{i,i}$ $(i \geqslant 1)$ | 11 | $pq$ | $s_{i+1,i+1}$ | $(0,0)$ |
| $s_{i,0}$ $(i \geqslant 2)$ | 00 | $qp$ | $s_{i,0}$ | $(0,0)$ |
| $s_{i,0}$ $(i \geqslant 3)$ | 01 | $q^2$ | $s_{i-1,0}$ | $(1,0)$ |
| $s_{i,0}$ $(i = 2)$ | 01 | $q^2$ | $s_0$ | $(2,0)$ |
| $s_{i,0}$ $(i \geqslant 2)$ | 10 | $p^2$ | $s_{i+1,0}$ | $(0,0)$ |
| $s_{i,0}$ $(i \geqslant 2)$ | 11 | $pq$ | $s_{i,0}$ | $(1,0)$ |

### 4.1. Система уравнений для вероятностей состояний

Для оценки вероятностей выигрыша каждой из групп необходимо сначала вычислить вероятности $p_i$ $(i = 0, 1, \ldots)$, $p_{i,0}$ $(j = 2, 3, \ldots)$ и $p_{i,i}$ $(i = 1, 2, \ldots)$ нахождения системы в каждом из состояний. Будем исходить из предположения, что соответствующая цепь Маркова является стационарной, т. е. эти вероятности не зависят от момента времени.

Вероятности нахождения системы в каждом из состояний должны удовлетворять следующим уравнениям:

$$p_0 = (p^2+q^2)\sum_{i\geqslant 1} p_{i,i} + qp_0 + pqp_1 + q^2(p_2+p_{2,0}); \tag{1}$$

$$p_i = pq\,p_i + p^2 p_{i-1}, \quad i \geqslant 1; \tag{2}$$

$$\begin{cases} p_{1,1} = qp\,p_{1,1} + pq\,p_0 + q^2 p_1, \\ p_{i,i} = qp\,p_{i,i} + pq\,p_{i-1,i-1} \quad (i \geqslant 2); \end{cases} \tag{3}$$

$$\begin{cases} p_{2,0} = qp\,p_{2,0} + q^2 p_{3,0} + pq\,p_{2,0} + pq\,p_2 + q^2 p_3, \\ p_{i,0} = qp\,p_{i,0} + p^2 p_{i-1,0} + q^2 p_{i+1,0} + pq\,p_{i,0} + pq\,p_i + q^2 p_{i+1} \quad (i \geqslant 3). \end{cases} \tag{4}$$

Наконец, общая сумма вероятностей всех состояний равна единице:

$$\sum_{i\geqslant 0} p_i + \sum_{i\geqslant 1} p_{i,i} + \sum_{i\geqslant 2} p_{i,0} = 1. \tag{5}$$

В дальнейшем найдём выражения для всех вероятностей через вероятность $p_1$ и значения $p, q$, $p < q$.

### 4.2. Вычисление вероятностей $p_i$

**Лемма 1.** Вероятности $p_i$ при $i \geqslant 0$ удовлетворяют соотношениям

$$p_{i+1} = \frac{p^2}{1-pq} p_i. \tag{6}$$

Для любого $k \geqslant 0$ имеем

$$\sum_{i\geqslant k} p_i = p_k \tfrac{1-pq}{q}. \tag{7}$$

***Доказательство.*** Первое соотношение равносильно уравнениям (2). Второе вытекает из первого и цепочки равенств $\sum\limits_{i\geqslant k} p_i = p_k \sum\limits_{i\geqslant 0} \left(\dfrac{p^2}{1-pq}\right)^i = p_k \dfrac{1-pq}{q}$. ■

В частности,

$$\sum_{i\geqslant 0} p_i = p_0 \frac{1-pq}{q} = p_1 \frac{(1-pq)^2}{p^2 q}. \tag{8}$$

**Следствие 1.** Имеют место следующие соотношения:

$$\begin{gathered} pq\,p_0 + q^2 p_1 = p_1 q/p, \\ pq\,p_i + q^2 p_{i+1} = p_1 \frac{p^{2i-1} q}{(1-pq)^i}, \quad i \geqslant 1. \end{gathered} \tag{9}$$

Оба соотношения вытекают из равенств (6).

### 4.3. Вычисление вероятностей $p_{i,i}$

**Лемма 2.** Вероятности $p_{i,i}$ при $i = 1, 2, \ldots$ удовлетворяют следующим соотношениям:

$$p_{1,1} = p_1 \frac{q}{p(1-pq)}, \qquad p_{i+1,i+1} = \frac{pq}{1-pq} p_{i,i}. \tag{10}$$

В частности,

$$\sum_{i\geqslant 1} p_{i,i} = p_1 \frac{q}{p(p^2+q^2)}. \tag{11}$$

***Доказательство.*** Уравнения подсистемы (3) можно переписать в более удобном виде:

$$\begin{cases} (1-pq)p_{1,1} = pq\,p_0 + q^2 p_1, \\ (1-pq)p_{i,i} = pq\,p_{i-1,i-1}, \quad i \geqslant 2. \end{cases} \tag{12}$$

Выражение для $p_{1,1}$ вытекает из первого уравнения этой системы и равенства (9). Второе соотношение вытекает из оставшихся уравнений системы (12). Третье соотношение является следствием первых двух и равенства $1 - 2pq = p^2 + q^2$:

$$\sum_{i\geqslant 1} p_{i,i} = p_{1,1} \sum_{i\geqslant 0} \left(\frac{pq}{1-pq}\right)^i = p_{1,1}\frac{1-pq}{1-2pq} = \frac{p_1 q}{p(p^2+q^2)}.$$

Лемма 2 доказана. ■

Для удобства обозначим через $\Sigma_0$ значение, определяемое выражением

$$p_1\Sigma_0 = \sum_{i\geqslant 1} p_{i,i} = p_{1,1} + p_{2,2} + \ldots$$

Далее нам потребуются также значения $\Sigma_1$ и $\Sigma_2$, определяемые равенствами

$$p_1\Sigma_1 = \sum_{i\geqslant 1}(i+1)p_{i,i} = 2p_{1,1} + 3p_{2,2} + \ldots,$$
$$p_1\Sigma_2 = \sum_{i\geqslant 1} i p_{i,i} = p_{1,1} + 2p_{2,2} + \ldots$$

**Лемма 3.** Значения $\Sigma_1$ и $\Sigma_2$ вычисляются по формулам

$$\Sigma_1 = \frac{q(2-3pq)}{p(1-2pq)^2}, \qquad \Sigma_2 = \frac{q(1-pq)}{p(1-2pq)^2}. \tag{13}$$

***Доказательство.*** Вычислим значение $\Sigma_1$. С учётом (10) имеем

$$p_1\Sigma_1 = p_{1,1}(2 + 3\frac{pq}{1-pq} + \ldots) = p_{1,1}\sum_{i\geqslant 1}(i+1)\left(\frac{pq}{1-pq}\right)^{i-1} =$$
$$= \frac{p_1}{p^2}\frac{pq}{1-pq}\sum_{i\geqslant 1}(i+1)\left(\frac{pq}{1-pq}\right)^{i-1} = \frac{p_1}{p^2}\sum_{i\geqslant 1}(i+1)\left(\frac{pq}{1-pq}\right)^{i} = \frac{p_1}{p^2}\left(\left(\frac{1-pq}{1-2pq}\right)^2 - 1\right).$$

Здесь использовано известное тождество

$$\sum_{i\geqslant 1} i x^{i-1} = \frac{1}{(1-x)^2}, \tag{14}$$

справедливое при $0 \leqslant x < 1$. Значит,

$$\Sigma_1 = \frac{1}{p^2}\left(\left(\frac{1-pq}{1-2pq}\right)^2 - 1\right) = \frac{1}{p^2}\left(\frac{2pq - 3p^2q^2}{(1-2pq)^2}\right) = \frac{q(2-3pq)}{p(1-2pq)^2}.$$

Аналогично с учётом (14) получаем

$$p_1\Sigma_2 = p_{1,1}(1 + 2\frac{pq}{1-pq} + \ldots) = p_{1,1}\sum_{i\geqslant 1} i\left(\frac{pq}{1-pq}\right)^{i-1} =$$
$$= \frac{p_1}{p^2}\frac{pq}{1-pq}\sum_{i\geqslant 1} i\left(\frac{pq}{1-pq}\right)^{i-1} = \frac{p_1}{p^2}\frac{pq}{1-pq}\left(\frac{1-pq}{1-2pq}\right)^2 = p_1\frac{q(1-pq)}{p(1-2pq)^2}.$$

Лемма 3 доказана. ■

4.4. Вычисление вероятностей $p_{i,0}$

**Утверждение 1.** Величины $p_{i,0}$, $i \geqslant 2$, вычисляются по следующим формулам:

$$p_{2,0} = p_1 \frac{p^3}{q^2(1-pq)}; \tag{15}$$

$$p_{i+1,0} = p_1 \left(\frac{p^2}{q^2}\right)^i \left(\frac{p+q^2}{1-pq} - \left(\frac{q^2}{1-pq}\right)^i\right), \quad i \geqslant 2. \tag{16}$$

***Доказательство.*** Из равенств (1), (7) и (11) получаем

$$p_2 + p_{2,0} = \frac{1}{q}\sum_{i\geqslant 2} p_i = p_1 \tfrac{p^2}{q^2}.$$

Поэтому

$$p_{2,0} = (p_2 + p_{2,0}) - p_2 = p_1\left(\frac{p^2}{q^2} - \frac{p^2}{1-pq}\right) = p_1 \frac{p^3}{q^2(1-pq)}. \tag{17}$$

Уравнения системы (4) можно переписать в более удобном виде:

$$\begin{cases} (1-2pq)p_{2,0} = q^2 p_{3,0} + pq\, p_2 + q^2 p_3, \\ (1-2pq)p_{i,0} = p^2 p_{i-1,0} + q^2 p_{i+1,0} + pq\, p_i + q^2 p_{i+1}, \quad i > 2, \end{cases}$$

или иначе

$$\begin{cases} (p^2+q^2)p_{2,0} = q^2 p_{3,0} + p_1 \dfrac{p^3 q}{(1-pq)^2}, \\ (p^2+q^2)p_{i,0} = p^2 p_{i-1,0} + q^2 p_{i+1,0} + p_1 \dfrac{p^{2i-1} q}{(1-pq)^i}, \quad i > 2. \end{cases}$$

Отсюда

$$\begin{cases} p_{2,0} - p_{3,0} = -\dfrac{p^2}{q^2} p_{2,0} + p_1 \dfrac{p^3}{q(1-pq)^2}, \\ p_{i,0} - p_{i+1,0} = \dfrac{p^2}{q^2}(p_{i-1,0} - p_{i,0}) + p_1 \dfrac{p^{2i-1}}{q(1-pq)^i}, \quad i > 2. \end{cases}$$

Обозначив через $y(i) = p_{i,0} - p_{i+1,0}$, получаем

$$\begin{cases} y(2) = -\dfrac{p^2}{q^2} p_{2,0} + p_1 \dfrac{p^3}{q(1-pq)^2}, \\ y(i) = \dfrac{p^2}{q^2} y(i-1) + p_1 \dfrac{p^{2i-1}}{q(1-pq)^i}, \quad i > 2. \end{cases}$$

Значит, при $i \geqslant 2$

$$y(i) = \frac{p^2}{q^2}\left(\dots \frac{p^2}{q^2}\left(-\frac{p^2}{q^2}p_{2,0} + p_1\frac{p^3}{q(1-pq)^2}\right) + p_1\frac{p^5}{q(1-pq)^3}\dots\right) + p_1\frac{p^{2i-1}}{q(1-pq)^i} =$$

$$= p_1 \sum_{j=2}^{i}\left(\frac{p^2}{q^2}\right)^{i-j} \frac{p^{2j-1}}{q(1-pq)^j} - \left(\frac{p^2}{q^2}\right)^{i-1} p_{2,0} = p_1 \sum_{j=2}^{i} \frac{p^{2i-1}}{q^{2i-2j+1}(1-pq)^j} - \left(\frac{p^2}{q^2}\right)^{i-1} p_{2,0} =$$

$$= p_1\left(\frac{p^{2i-1}}{q^{2i+1}} \sum_{j\geqslant 2}^{i}\left(\frac{q^2}{1-pq}\right)^j - \left(\frac{p^2}{q^2}\right)^{i-1} p_{2,0}/p_1\right) =$$

$$= p_1 \left( \frac{p^{2i-1}}{q^{2i+1}} \left( \frac{q^2}{1-pq} \right)^2 \sum_{j \geqslant 0}^{i-2} \left( \frac{q^2}{1-pq} \right)^j - \left( \frac{p^2}{q^2} \right)^{i-1} p_{2,0}/p_1 \right) =$$

$$= p_1 \left( \frac{p^{2i-1}}{q^{2i+1}} \left( \frac{q^2}{1-pq} \right)^2 \frac{1-pq}{p} \left( 1 - \left( \frac{q^2}{1-pq} \right)^{i-1} \right) - \left( \frac{p^2}{q^2} \right)^{i-1} p_{2,0}/p_1 \right) =$$

$$= p_1 \left( \left( \frac{p^2}{q^2} \right)^{i-1} \frac{q}{1-pq} \left( 1 - \left( \frac{q^2}{1-pq} \right)^{i-1} \right) - \left( \frac{p^2}{q^2} \right)^{i-1} p_{2,0}/p_1 \right) =$$

$$= p_1 \left( \left( \frac{q}{1-pq} - p_{2,0}/p_1 \right) \left( \frac{p^2}{q^2} \right)^{i-1} - \frac{q}{1-pq} \left( \frac{q^2}{1-pq} \right)^{i-1} \right) =$$

$$= p_1 \left( \frac{q^3-p^3}{q^2(1-pq)} \left( \frac{p^2}{q^2} \right)^{i-1} - \frac{q}{(1-pq)} \left( \frac{p^2}{1-pq} \right)^{i-1} \right).$$

Таким образом,

$$p_{i+1,0} = p_{i,0} - p_1 \left( \frac{q^3-p^3}{q^2(1-pq)} \left( \frac{p^2}{q^2} \right)^{i-1} - \frac{q}{(1-pq)} \left( \frac{p^2}{1-pq} \right)^{i-1} \right) =$$

$$= p_{2,0} - p_1 \sum_{j=2}^{i} \left( \frac{q^3-p^3}{q^2(1-pq)} \left( \frac{p^2}{q^2} \right)^{j-1} - \frac{q}{(1-pq)} \left( \frac{p^2}{1-pq} \right)^{j-1} \right) =$$

$$= p_1 \frac{p^3}{q^2(1-pq)} - p_1 \left( \frac{q^3-p^3}{q^2(1-pq)} \sum_{j=2}^{i} \left( \frac{p^2}{q^2} \right)^{j-1} - \frac{q}{(1-pq)} \sum_{j=2}^{i} \left( \frac{p^2}{1-pq} \right)^{j-1} \right) =$$

$$= p_1 \frac{p^3}{q^2(1-pq)} - p_1 \left( \frac{q^3-p^3}{q^2(1-pq)} \frac{p^2}{q-p} \left( 1 - \left( \frac{p^2}{q^2} \right)^{i-1} \right) - \frac{q}{(1-pq)} \frac{p^2}{q} \left( 1 - \left( \frac{p^2}{1-pq} \right)^{i-1} \right) \right) =$$

$$= p_1 \left( \frac{q^3-p^3}{(1-pq)} \frac{1}{q-p} \left( \frac{p^2}{q^2} \right)^i - \left( \frac{p^2}{1-pq} \right)^i \right) = p_1 \left( \frac{p^2}{q^2} \right)^i \left( \frac{p+q^2}{1-pq} - \left( \frac{q^2}{1-pq} \right)^i \right).$$

Утверждение 1 доказано. ■

**Следствие 2.** Значение $\Sigma_3$, определяемое выражением

$$\sum_{i \geqslant 2} p_{i,0} = p_1 \Sigma_3,$$

вычисляется по формуле

$$\Sigma_3 = \frac{p^3}{q(q-p)}. \tag{18}$$

***Доказательство.*** Действительно, из (16) имеем

$$\sum_{i \geqslant 3} p_{i,0} = \sum_{i \geqslant 2} p_{i+1,0} = p_1 \sum_{i \geqslant 2} \left( \frac{p^2}{q^2} \right)^i \left( \frac{p+q^2}{1-pq} - \left( \frac{q^2}{1-pq} \right)^i \right) =$$

$$= p_1 \left( \frac{p+q^2}{1-pq} \sum_{i \geqslant 2} \left( \frac{p^2}{q^2} \right)^i - \sum_{i \geqslant 2} \left( \frac{p^2}{1-pq} \right)^i \right) = p_1 \left( \frac{p+q^2}{1-pq} \left( \frac{p^2}{q^2} \right)^2 \frac{q^2}{q-p} - \left( \frac{p^2}{1-pq} \right)^2 \frac{1-pq}{q} \right) =$$

$$= p_1 \frac{p^4}{q^4(1-pq)} \left( \frac{q^2(p+q^2)}{(q-p)} - q^3 \right) = p_1 \frac{p^5(1+q)}{q^2(1-pq)(q-p)}.$$

Отсюда с учётом (15) получаем

$$\sum_{i\geqslant 2} p_{i,0} = p_{2,0} + \sum_{i\geqslant 3} p_{i,0} = p_1 \frac{p^3}{q^2(1-pq)} \left(1 + \frac{p^2(1+q)}{q-p}\right) = p_1 \frac{p^3}{q(q-p)}.$$

Следствие 2 доказано. ■

### 4.5. Вычисление вероятности $p_1$

Из равенства (5) с учётом (7), (11) и (18) получаем выражение для вычисления вероятности $p_1$:

$$p_1 \left(\frac{(1-pq)^2}{p^2 q} + \frac{q}{p(p^2+q^2)} + \frac{p^3}{q(q-p)}\right) = 1.$$

### 4.6. Вычисление $R$

Оценим выигрыш, получаемый выделенной группой участников. Вознаграждение первой группы с учётом (9), (7), (17), (18) и (13) равно

$$r_0 = 2pq\,p_1 + q^2\,p_2 + (pq+q^2)\sum_{i\geqslant 2} p_i + q^2 p_{2,0} + (qp+q^2)\sum_{i\geqslant 2} p_{i,0} + p^2 \sum_{i\geqslant 1}(i+1)p_{i,i} =$$

$$= p_1 \left(2pq + \frac{p^2q^2}{1-pq} + p^2 + \frac{p^3}{1-pq} + q\Sigma_3 + p^2\Sigma_1\right).$$

Для второй группы вознаграждение равно

$$r_1 = q^2 \left(p_0 + \sum_{i\geqslant 1}(i+1)p_{i,i}\right) = p_1 q^2 \left(\frac{1-pq}{p^2} + \Sigma_1\right).$$

Доля первой группы в общей сумме вознаграждения по результатам майнинга оценивается выражением $R = r_0/(r_0+r_1)$; заметим, что оно не зависит от $p_1$. На рис. 3 приведены результаты вычислений величины выигрыша $R$ при честном и корыстном майнинге в зависимости от вероятности $p$. Из графика видно, что корыстная группа получает преимущество по сравнению с честным майнингом при $p > 0{,}429$, причём её вознаграждение превышает вознаграждение остальной группы при $p > 0{,}454$.

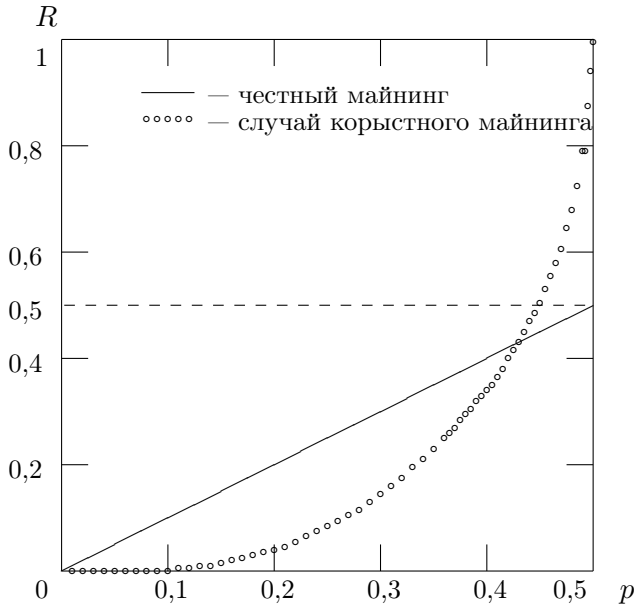



Рис. 3. График зависимости величины выигрыша $R$ при честном и корыстном майнинге от величины вероятности $p$

## 5. Использование предположения о подключении части участников из второй группы

Теперь по аналогии с [1] рассмотрим ситуацию, когда в случае разветвления блокчейна на две цепочки часть майнеров, не входящих в выделенную группу, подключится к поиску продолжения цепочки, найденной выделенной группой участников. Пусть доля таких участников в общей вычислительной мощности второй группы участников оценивается величиной $\gamma$, $0 \leqslant \gamma \leqslant 1$.

Для анализа этой ситуации будем, как и раньше, рассматривать вероятностную модель, включающую две группы участников, осуществляющих майнинг с вероятностями успеха $p$ и $q$. В тех случаях, когда для группы остальных участников имеется выбор того, для какой из цепочек строить продолжение, будем предполагать, что вероятность успешного подбора продолжения для цепочки, содержащей блоки, найденные группой корыстных участников, равна $\gamma q$, а для второй цепочки в разветвлении блокчейна она равна $(1-\gamma)q$.

Заметим, что модель с независимым поведением каждой из подгрупп оказывается слишком сложной для исследования, так как при этом появляется возможность таких переходов в новое состояние, когда все три участника добиваются успеха. Это приводит к необходимости рассмотрения тройных, четверных и т. д. разветвлений блокчейна, что приводит к дополнительному увеличению числа участников и значительно усложняет вид графа переходов и всей марковской модели.

С другой стороны, в данном случае обе подгруппы используют честную стратегию, решая свои задачи в открытую. Для простоты предположим, что они так же, как и корыстная группа, находятся в одном пуле, управляющем распределением работ и сбором результатов. Поэтому естественно считать, что все члены группы прекращают перебор в момент успешного подбора блока одной из подгрупп. Следовательно, здесь возможно использование одной случайной величины с тремя исходами: ни одна из подгрупп не нашла варианта (этот вариант остаётся, так как $q < 1$), успешна первая или вторая подгруппа.

Таким образом, для тех состояний, которые соответствуют разветвлению блокчейна, должно быть не четыре, а шесть вариантов перехода в другие состояния: (два варианта для корыстной группы)$\times$(три варианта для группы остальных участников). Такому порядку проведения майнинга соответствует модель с двумя группами: одна — это корыстная группа, работающая в соответствии с алгоритмом корыстного майнинга, вторая — группа остальных майнеров, выдающая результатом подбора продолжение либо первой, либо второй цепочки. Её работа моделируется случайной величиной, принимающей три значения 0, 1, 2 с вероятностями $p$, $q\gamma$, $q(1-\gamma)$ соответственно.

Соберём в одну таблицу варианты перехода для всех состояний (табл. 3). Первая часть таблицы такая же, как и в табл. 2, так как она соответствует состояниям, в которых корыстная группа скрывает свою часть цепочки, и поэтому вторая группа целиком работает на продолжение опубликованной части цепочки. Отличие имеется только для состояний, в которых имеется разветвление. Из табл. 3 видно, что при $0 < \gamma < 1$ вероятность выигрыша корыстной группы должна увеличиться, так как в графе появляются дополнительные переходы, приносящие успех первой группе. Модифицированный граф переходов, моделирующий этот случай, приведён на рис. 4.

Таблица 3

**Переходы модифицированного графа**

| Исходное состояние | Метка ребра | Верояность перехода | Следующее состояние | Выигрыш обеих групп |
|---|---|---|---|---|
| $s_i$ $(i \geqslant 0)$ | 00 | $qp$ | $s_i$ | $(0,0)$ |
| $s_0$ | 01 | $q^2$ | $s_0$ | $(0,1)$ |
| $s_1$ | 01 | $q^2$ | $s_{1,1}$ | $(0,0)$ |
| $s_2$ | 01 | $q^2$ | $s_0$ | $(2,0)$ |
| $s_i$ $(i \geqslant 3)$ | 01 | $q^2$ | $s_{i-1,0}$ | $(1,0)$ |
| $s_i$ $(i \geqslant 0)$ | 10 | $p^2$ | $s_{i+1}$ | $(0,0)$ |
| $s_0$ | 11 | $pq$ | $s_{1,1}$ | $(0,0)$ |
| $s_1$ | 11 | $pq$ | $s_0$ | $(2,0)$ |
| $s_i$ $(i \geqslant 2)$ | 11 | $pq$ | $s_{i,0}$ | $(1,0)$ |
| $s_{i,i}$ | 00 | $qp$ | $s_{i,i}$ | $(0,0)$ |
| $s_{i,i}$ | 01 | $q^2\gamma$ | $s_0$ | $(i,1)$ |
| $s_{i,i}$ | 02 | $q^2(1-\gamma)$ | $s_0$ | $(0,i+1)$ |
| $s_{i,i}$ | 10 | $p^2$ | $s_0$ | $(i+1,0)$ |
| $s_{i,i}$ | 11 | $pq\gamma$ | $s_{1,1}$ | $(i,0)$ |
| $s_{i,i}$ | 12 | $pq(1-\gamma)$ | $s_{i+1,i+1}$ | $(0,0)$ |
| $s_{i,0}$ | 00 | $qp$ | $s_{i,0}$ | $(0,0)$ |
| $s_{2,0}$ | 01 | $q^2\gamma$ | $s_0$ | $(2,0)$ |
| $s_{i,0}$ $(i>2)$ | 01 | $q^2\gamma$ | $s_{i-1,0}$ | $(1,0)$ |
| $s_{2,0}$ | 02 | $q^2(1-\gamma)$ | $s_0$ | $(2,0)$ |
| $s_{i,0}$ $(i>2)$ | 02 | $q^2(1-\gamma)$ | $s_{i-1,0}$ | $(1,0)$ |
| $s_{i,0}$ | 10 | $p^2$ | $s_{i+1,0}$ | $(0,0)$ |
| $s_{i,0}$ | 11 | $pq\gamma$ | $s_{i,0}$ | $(1,0)$ |
| $s_{i,0}$ | 12 | $pq(1-\gamma)$ | $s_{i,0}$ | $(1,0)$ |

Система уравнений для вероятностей состояний теперь имеет следующий вид:

$$p_0 = (p^2+q^2)\sum_{i\geqslant 1} p_{i,i} + qp_0 + pqp_1 + q^2\left(p_2 + p_{2,0}\right); \tag{19}$$

$$p_i = pq\,p_i + p^2 p_{i-1}, \quad i \geqslant 1; \tag{20}$$

$$\begin{cases} p_{1,1} = qp\,p_{1,1} + pq\,p_0 + q^2 p_1 + pq\gamma \sum\limits_{j\geqslant 1} p_{j,j}, \\ p_{i,i} = pq\,p_{i,i} + pq(1-\gamma)\,p_{i-1,i-1}, \quad i \geqslant 2; \end{cases} \tag{21}$$

$$\begin{cases} p_{2,0} = qp\,p_{2,0} + q^2 p_{3,0} + pq\,p_{2,0} + pq\,p_2 + q^2 p_3, \\ p_{i,0} = qp\,p_{i,0} + p^2 p_{i-1,0} + q^2 p_{i+1,0} + pq\,p_{i,0} + pq\,p_i + q^2 p_{i+1}, \quad i \geqslant 3. \end{cases} \tag{22}$$

В дальнейшем поступим аналогично предыдущему случаю и найдём выражения для всех вероятностей через вероятность $p_1$ и значения параметров $p, q$ и $\gamma$.

Заметим, что уравнения (19), (20) и (22) совпадают с (1), (2) и (4) соответственно. Поэтому для вероятностей $p_i$, $i \geqslant 0$, справедливы формулы (6), (7) и (8). Отличие имеется только в первом уравнении для вероятности $p_{11}$ в подсистемах (3) и (21).

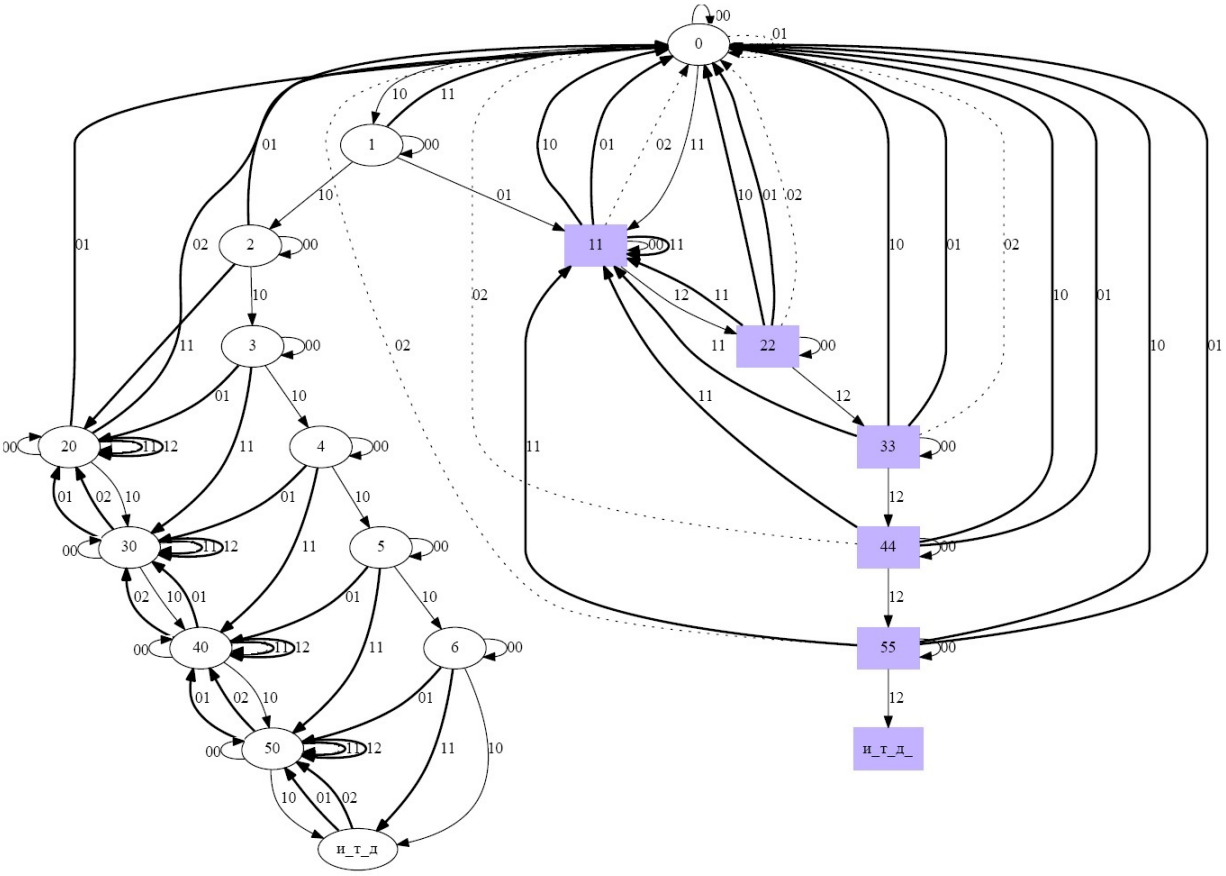



Рис. 4. Автоматная модель при $0 \leqslant \gamma \leqslant 1$

### 5.1. Вычисление вероятностей $p_{i,i}$

**Лемма 4.** Справедливы следующие равенства:

$$p_{i+1,i+1} = \frac{pq(1-\gamma)}{1-pq} p_{i,i}, \qquad i = 1, 2, \ldots; \tag{23}$$

$$p_{1,1} = p_1 \frac{q(1-pq(2-\gamma))}{p(1-pq)(1-2pq)}; \tag{24}$$

$$\sum_{i \geqslant 1} p_{i,i} = p_{1,1} \frac{1-pq}{1-pq(2-\gamma)} = p_1 \frac{q}{p(1-2pq)}. \tag{25}$$

***Доказательство.*** Преобразуем выражения (21):

$$\begin{cases} (1-pq)p_{1,1} = pq\, p_0 + q^2 p_1 + pq\gamma \sum\limits_{j \geqslant 1} p_{j,j}, \\ (1-pq)p_{i,i} = pq(1-\gamma)\, p_{i-1,i-1}, \quad i \geqslant 2. \end{cases} \tag{26}$$

Равенства (23) получаются из второго уравнения системы (26), поэтому

$$\sum_{i \geqslant 1} p_{i,i} = p_{1,1} \sum_{i \geqslant 0} \left( \frac{pq(1-\gamma)}{1-pq} \right)^i = p_{1,1} \frac{1-pq}{1-pq(2-\gamma)}. \tag{27}$$

Для нахождения выражения для вероятности $p_{1,1}$ необходимо вычислить правую сумму в правой части первого уравнения системы (26). Используя равенство (9), имеем

$$p_{1,1} = p_1 \frac{q}{p(1-pq)} + \frac{pq\gamma \sum\limits_{j\geqslant 1} p_{j,j}}{1-pq}. \tag{28}$$

Из (27) и (28) получаем равенство

$$p_{1,1}\left(1 - \frac{pq\gamma}{1-pq(2-\gamma)}\right) = p_1 \frac{q}{p(1-pq)},$$

что равносильно (24). Подставляя выражение (24) в (27), получаем требуемое выражение (25) для суммы. ■

Рассмотрим теперь, как изменятся значения сумм $\Sigma_0$, $\Sigma_1$ и $\Sigma_2$. Выражение (25) совпадает с (11), поэтому значение $\Sigma_0$ остается неизменным.

**Лемма 5.** Значения $\Sigma_1$ и $\Sigma_2$ вычисляются по формулам

$$\Sigma_1 = \frac{q(2-pq(3-\gamma))}{p(1-2pq)(1-pq(2-\gamma))}, \qquad \Sigma_2 = \frac{q(1-pq)}{p(1-2pq)(1-pq(2-\gamma))}. \tag{29}$$

***Доказательство.*** В силу тождества (14) с учётом (24) справедливы формулы

$$p_1\Sigma_2 = \sum_{i\geqslant 1} i p_{i,i} = p_{1,1} \sum_{i\geqslant 1} i \left(\frac{pq(1-\gamma)}{1-pq}\right)^{i-1} = p_{1,1} \frac{1}{\left(1 - \dfrac{pq(1-\gamma)}{1-pq}\right)^2} =$$

$$= p_{1,1}\left(\frac{1-pq}{(1-pq(2-\gamma)}\right)^2 = p_1 \frac{q(1-pq)}{p(1-2pq)(1-pq(2-\gamma))}.$$

Отсюда

$$p_1\Sigma_1 = p_1(\Sigma_2 + \Sigma_0) = p_{1,1}\frac{(1-pq)(2-pq(3-\gamma))}{(1-pq(2-\gamma))^2} = p_1 \frac{q(2-pq(3-\gamma))}{p(1-2pq)(1-pq(2-\gamma))}.$$

Лемма доказана. ■

### 5.2. Вычисление вероятностей $p_{i,0}$

Поскольку выражение (25) не зависит от $\gamma$, значение вероятности $p_{2,0}$, вычисляемое с помощью уравнения (19), остаётся неизменным и определяется выражением (15). Поэтому значения вероятностей $p_{i,0}$, $i \geqslant 3$, также не меняются и определяются выражениями (16).

### 5.3. Вычисление вознаграждения для обеих групп

Перейдём к оценке величины $R = r_0/(r_0 + r_1)$ доли вознаграждения корыстной группы в общей сумме вознаграждения, полученной при применении описанной стратегии майнинга. Вознаграждение первой группы в этом случае определяется как

$$r_0 = 2pq\,p_1 + q^2\,p_2 + (pq+q^2)\sum_{i\geqslant 2} p_i + q^2 p_{2,0} + (qp+q^2)\sum_{i\geqslant 2} p_{i,0} + p^2 \sum_{i\geqslant 1}(i+1)p_{ii} +$$

$$+ (q^2+pq)\gamma \sum_{i\geqslant 1} i p_{ii} = p_1\left(2pq + \frac{p^2q^2}{1-pq} + p^2 + \frac{p^3}{1-pq} + q\Sigma_3 + p^2\Sigma_1 + q\gamma\Sigma_2\right).$$

Для второй группы вознаграждение равно

$$r_1 = p_1 q^2 \left( \frac{1 - pq}{p^2} + (1 - \gamma)\Sigma_1 + \gamma\Sigma_0 \right).$$

Значения сумм $\Sigma_0$, $\Sigma_1$ и $\Sigma_2$ вычисляются с помощью тождеств (11) и (29).

Приведённые формулы позволяют вычислить значение доли $R$ при произвольных значениях параметров $0 \leqslant p < 1/2$ и $0 \leqslant \gamma \leqslant 1$. Результаты вычислений приведены на рис. 5–7. На рис. 5 показан график зависимости от параметра $\gamma$ минимального значения вероятности $p$, при котором впервые выполняется условие $R > 1/2$. Вычисления показывают, что выигрыш корыстной группы при соответствующем значении $\gamma$ превышает выигрыш остальной группы при значениях вероятности $p$ в пределах

$$0{,}358 \leqslant p \leqslant 0{,}449.$$

Наибольшее значение достигается при $\gamma = 0$, а наименьшее — при $\gamma = 1$. На рис. 6 показан аналогичный график зависимости от параметра $\gamma$ минимального значения вероятности $p$, при котором впервые выполняется условие $R > p$. В данном случае получаем, что выигрыш корыстной группы превышает при соответствующем значении $\gamma$ выигрыш, полученный ими при честном выполнении блокчейн протокола, для значений вероятности $p$ в пределах

$$0 < p \leqslant 0{,}429.$$

В работе [1] этот интервал имеет вид $0 < p \leqslant 0{,}333$.

На рис. 7 приведён график величины доли вознаграждения $R$ при честном и корыстном майнинге в зависимости от величины вероятности $p$ для трёх значений параметра $\gamma$ (0, 0,5 и 1).

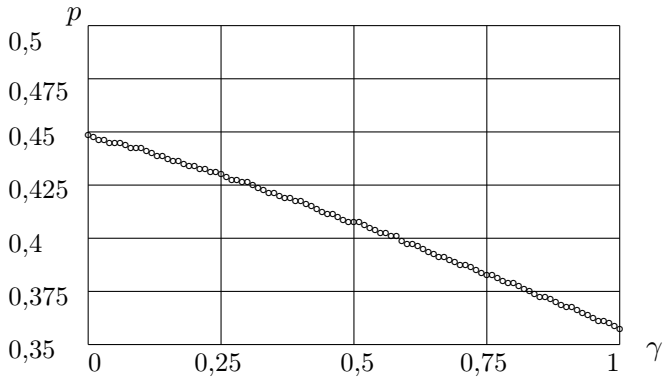



Рис. 5. График зависимости значения вероятности $p$, при котором впервые выполняется условие $R > 1/2$, от параметра $\gamma$

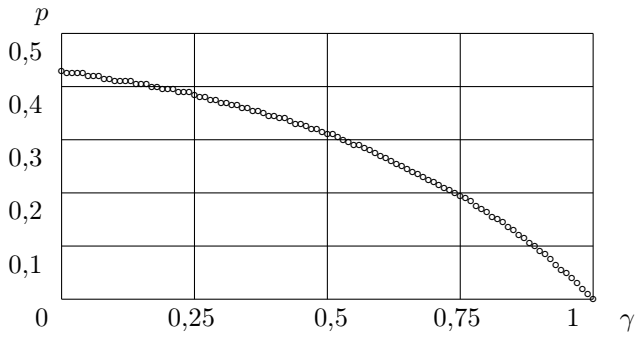



Рис. 6. График зависимости значения вероятности $p$, при котором впервые выполняется условие $R > p$, от параметра $\gamma$

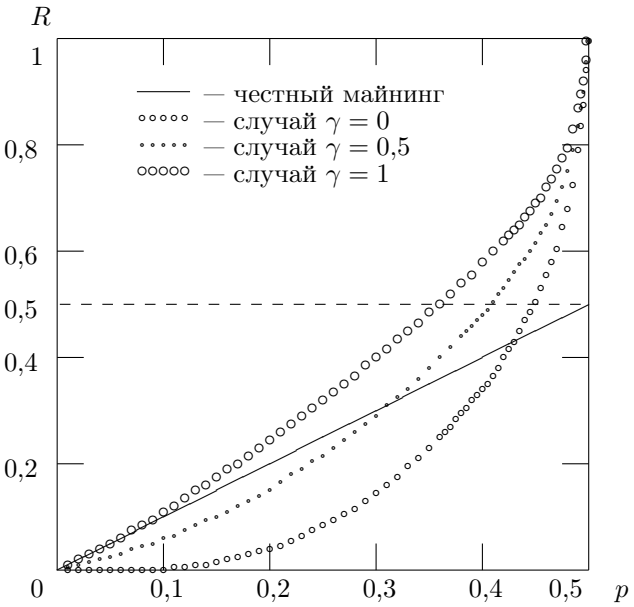



Рис. 7. Графики зависимости величины выигрыша $R$ при корыстном майнинге в зависимости от величины вероятности $p$ и параметра $\gamma$

Автор выражает благодарность рецензенту за внимательное прочтение рукописи и многочисленные полезные замечания.

## ЛИТЕРАТУРА

1. *Ittay E. and Emin G. S.* Majority is Not Enough: Bitcoin Mining is Vulnerable. arXiv:1311.0243. 2013. http://arxiv.org/abs/1311.0243.
2. *Ittay E. and Emin G. S.* Majority is not enough: bitcoin mining is vulnerable // Financial Cryptography and Data Security: 18th Intern. Conf. Christ Church, Barbados, March 3–7, 2014. P. 436–454.
3. *Ittay E. and Emin G. S.* Majority is not enough: bitcoin mining is vulnerable // Commun. ACM. 2018. V. 61. No. 7. P. 95–102. https://doi.org/10.1145/3212998.

REFERENCES

1. *Ittay E. and Emin G. S.* Majority is Not Enough: Bitcoin Mining is Vulnerable. arXiv:1311.0243. 2013. `http://arxiv.org/abs/1311.0243`.
2. *Ittay E. and Emin G. S.* Majority is not enough: bitcoin mining is vulnerable. Financial Cryptography and Data Security: 18th Intern. Conf. Christ Church, Barbados, March 3–7, 2014, pp. 436–454.
3. *Ittay E. and Emin G. S.* Majority is not enough: bitcoin mining is vulnerable. Commun. ACM, 2018, vol. 61, no. 7, pp. 95–102. `https://doi.org/10.1145/3212998`.

# МАТЕМАТИЧЕСКИЕ ОСНОВЫ КОМПЬЮТЕРНОЙ БЕЗОПАСНОСТИ

УДК 519.7

## ЗАЩИЩЁННОЕ ХРАНЕНИЕ ДАННЫХ И ПОЛНОДИСКОВОЕ ШИФРОВАНИЕ

Е. К. Алексеев, Л. Р. Ахметзянова, А. А. Бабуева, С. В. Смышляев

*ООО «КРИПТО-ПРО», г. Москва, Россия*

Рассматривается задача защищённого хранения данных в современных операционных системах. Обсуждаются различные подходы к её решению, реализуемые программным образом. Основное внимание уделяется системам полнодискового шифрования как наиболее универсальной из технологий защиты хранимых данных. Перечисляются эксплуатационные и криптографические свойства, которые необходимо учитывать при их проектировании и сравнении. Описываются некоторые типичные сценарии использования таких криптосистем. Результаты работы могут быть использованы при синтезе и сравнении систем полнодискового шифрования.

**Ключевые слова:** *модели и методы защиты информации, защита хранимых данных.*

DOI 10.17223/20710410/49/6

## DATA STORAGE SECURITY AND FULL DISK ENCRYPTION

E. K. Alekseev, L. R. Akhmetzyanova, A. A. Babueva, S. V. Smyshlyaev

*CryptoPro, Moscow, Russia*

**E-mail:** alekseev@cryptopro.ru, lah@cryptopro.ru, babueva@cryptopro.ru, svs@cryptopro.ru

In the paper, a systematic description of the process of providing the security of data storage in modern operating systems is presented. The advantages of Full Disk Encryption (FDE) modules as compared with the other ways to security of this data storage are considered and explained. For most of modern FDE modules, there are four stages of work, namely: setup — initial data encryption, mounting — unfolding the key system in OS memory, session — reading and writing data using the FDE module (interaction of the file system with the hard disk driver), and unmounting — carrying out operations for ensuring purposeful properties of security and finishing work with the FDE module. These stages are introduced for the operating FDE module, including possible disrepairs, which are also systematized and considered in details. Performance characteristics that are important for synthesis and analysis are listed. Also, their target protective properties are studied in detail, the relationship between the problems of ensuring the confidentiality and integrity of data storage is shown and substantiated. New variants of these security properies are introduced so

that they can become a guideline in the creation of FDE modules and a possible trade-off between performance and security. Some typical scenarios of using such systems are described.

**Keywords:** *models and methods in information security, data storage security.*

## Введение

При разработке и применении средств криптографической защиты информации и, в том числе, средств электронной подписи принято учитывать требования не только к безопасности реализации самих криптографических преобразований, но и в целом к защите систем, на которых происходит обработка данных. Начиная с определённых классов защиты, требуется учитывать возможность противника совершать атаки из пределов контролируемой зоны [1, п. А.4.3], в том числе при наличии у противника доступа к атакуемой системе в качестве легитимного пользователя [2, п. 15 Приложения 1].

Таким образом, в ряде случаев как обрабатываемые пользователем данные, так и служебные файлы криптосредства необходимо защищать от атак со стороны других пользователей той же системы. Основным классом механизмов защиты, противодействующих подобным атакам, являются системы разграничения доступа, в том числе встроенные в операционные системы [1, п. 5.5.4]. Действительно, корректно настроенная система разграничения доступа не позволит противнику, работающему в системе, получить доступ как к данным атакуемого пользователя, так и к служебным файлам (например, базе настроек или журналу аудита).

Однако в реальной жизни именно вокруг уязвимостей в системах разграничения доступа и их настройках ведутся ожесточённые бои между разработчиками (операционных систем, а также прикладного и системного программного обеспечения) и злоумышленниками: иллюстрацией этого тезиса может являться количество уязвимостей в базах CVE, которые можно найти по ключевым словам «access control» и «privileges». Кроме того, зачастую заточенные под обеспечение максимального уровня безопасности настройки разграничения доступа существенно усложняют использование информационных систем простыми пользователями, а их обход для получения прямого доступа к хранимым данным остаётся возможным с помощью посекторного чтения жёсткого диска после извлечения его из компьютера. При этом в случае совместного использования пользователями одной виртуальной машины никакие меры по разграничению доступа в рамках операционной системы не помогут против злоумышленника, владеющего доступом к физической машине.

По этим причинам крайне важной является задача обеспечения конфиденциальности и целостности хранимых данных, чтобы даже злоумышленник с частичным или полным доступом к диску (но не к используемой ключевой информации, которая, как правило, хранится на отчуждаемых носителях) не смог ни прочитать защищаемые данные, ни внести в них несанкционированные изменения.

Настоящая работа посвящена системам полнодискового шифрования. Рассматриваются эксплуатационные и криптографические особенности построения, применения и анализа данного класса криптосистем. Насколько известно авторам, в отечественной литературе данная тема подробно не рассматривалась, поэтому об указанных криптосистемах можно найти лишь краткие упоминания [3]. При этом в зарубежной литературе полнодисковому шифрованию уделено достаточно много внимания. На сегодняшний день по числу рассматриваемых вопросов выделяются диссертации [4, 5],

содержащие обширный список полезных ссылок. Отметим, что большинство работ посвящено конкретным схемам полнодискового шифрования, при этом работ, посвящённых определению и формализации целевых для таких схем свойств безопасности, гораздо меньше (см., например, [6, 7]).

В работе приводится систематизированное описание процесса хранения данных в современных операционных системах, поясняется преимущество систем полнодискового шифрования перед другими подходами к защите этих данных. Вводятся этапы работы систем полнодискового шифрования, в том числе с учётом возможных сбоев, которые также систематизированы и подробно рассмотрены. Перечисляются важные с точки зрения синтеза и анализа эксплуатационные характеристики таких систем. Подробно рассмотрены целевые свойства безопасности, указана и обоснована связь задач обеспечения конфиденциальности и целостности хранимых данных. Введены новые градации этих свойств безопасности, которые могут послужить ориентиром при создании систем полнодискового шифрования и возможным компромиссом между производительностью и безопасностью. Перечислены некоторые типичные сценарии использования таких систем.

## 1. Хранение и защита данных в ОС

При сохранении данных на диск в большинстве современных операционных систем выделяются следующие уровни, на которых могут осуществляться сопутствующие этому процессу операции:

— уровень прикладных программных компонентов;
— уровень файловой системы;
— уровень драйвера диска;
— уровень контроллера жесткого диска.

Отметим, что потребность в сохранении данных может возникнуть не только на уровне прикладных компонентов, но и, например, на уровне файловой системы (пример таких данных — время создания файла). В этом случае такие данные обычно называют служебными данными соответствующего уровня, а операции по их сохранению осуществляются начиная с того уровня, где они появились.

Взаимодействие между прикладными программными компонентами и контроллером диска осуществляется следующим образом (схематически данный процесс представлен на рис. 1):

— Прикладная компонента при работе с диском оперирует интерфейсом, предоставляемым файловой системой. При этом информация имеет древовидную структуру, которая описывается в терминах директорий и файлов.
— Файловая система посылает запрос на чтение или запись данных диска с помощью интерфейса, предоставляемого драйвером диска. При этом единицей чтения и записи данных является логический сектор — байтовый массив фиксированного размера (как правило, этот размер кратен 512).
— Драйвер диска обращается для записи или чтения данных к контроллеру жёсткого диска. На этом этапе взаимодействие осуществляется в терминах, максимально приближенных к физической структуре диска (например, дорожек и физических секторов).

Вопрос защиты хранимых данных имеет ярко выраженный прикладной характер, поэтому при рассмотрении подходов к обеспечению безопасности данных необходимо уделять особое внимание их эксплуатационным свойствам. Для определённости

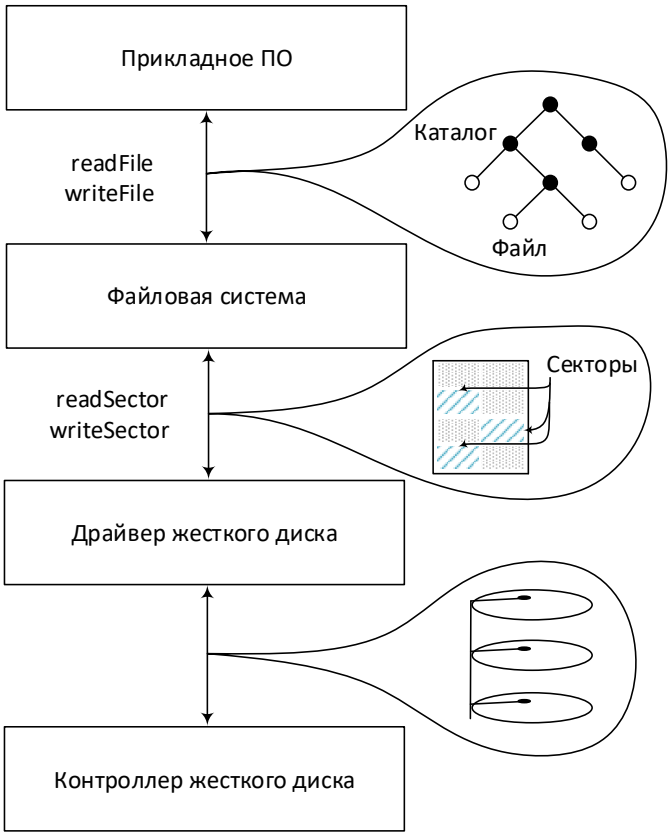



Рис. 1. Порядок взаимодействия файловой системы и жесткого диска

будем считать, что каждый подход реализуется некоторым модулем, функционирующим в рамках ОС. Учитывая иерархичность и многокомпонентность системы обработки данных, особого внимания заслуживает универсальность подхода, которая тем выше, чем больше число компонент, для которых модуль защиты может обеспечить свойства безопасности, и чем меньше дополнительных действий (например, изменений в программном коде) для этого необходимо осуществить.

Универсальность подхода во многом определяется положением модуля защиты в общей архитектуре системы, т. е. тем, где он расположен с точки зрения перечисленных уровней взаимодействия. Модуль защиты данных гипотетически может функционировать на любом из этих уровней, причём он может быть как встроен в штатную программную компоненту некоторого уровня, так и располагаться между уровнями, работая «прозрачно» для вышестоящих компонент. Под прозрачностью понимается то, что модуль полностью повторяет интерфейс компоненты, находящейся непосредственно под ним. За счёт этого модулю удаётся обеспечить защиту для всех вышестоящих компонент без необходимости внесения в них каких-либо изменений (иногда говорят, что вышестоящие компоненты даже «не знают» о том, что для сохраняемых ими данных обеспечивается защита). Стоит отметить, что модуль защиты не в состоянии обеспечить какую-либо безопасность служебных данных любых нижестоящих компонент.

Отметим, что обычно для реализации защиты ниже уровня драйвера диска используются аппаратные решения. Примером аппаратной реализации защиты храни-

мых данных является SSD-диск со встроенным шифрованием данных «Integral Crypto SSD» (другое название — «Integral Memory Crypto Hard Drive») [8]. Мы ограничимся рассмотрением только программных модулей защиты данных и не будем рассматривать варианты, когда защита осуществляется на уровне ниже драйвера диска.

Приведём примеры осуществления защиты хранимых данных на разных уровнях взаимодействия.

— Функция шифрования документов, встроенная в текстовый редактор Microsoft Word. Модуль защиты интегрирован в конкретную прикладную программную компоненту и не предназначен для защиты данных других приложений.
— Модуль защиты файлов AxCrypt является компонентой прикладного уровня и может обеспечить защиту любого файла, вне зависимости от того, какой прикладной компонентой он был создан. Однако для его применения необходимо каждый раз дополнительно указывать, какой файл должен быть защищён (или отдельно производить встраивание в случае предоставления соответствующего программного интерфейса).
— Функции защиты данных, реализованные в файловой системе NTFS, обычно объединяются под общим названием EFS (Encrypting File System). Эти функции формируют модуль защиты, который, по сути, реализован внутри файловой системы. Обеспечивая прозрачную защиту данных любых компонент прикладного уровня, EFS при этом не может защитить данные, обрабатываемые другими файловыми системами.
— Программный модуль VeraCrypt обеспечивает защиту всех секторов жёсткого диска. Он функционирует между файловой системой и драйвером жёсткого диска.

Заметим, что драйвер диска, если и порождает в процессе работы служебные данные, то они имеют крайне незначительный объём, а необходимость в их защите минимальна. Поэтому подход, предполагающий реализацию модуля защиты между драйвером диска и файловой системой, представляется наиболее универсальным. Например, в его рамках можно обеспечить защиту такой зачастую критически важной информации, как топология файловой системы, имена хранимых файлов, их размеры и даты модификации. Далее рассматриваются свойства модулей защиты именно такого типа.

## 2. Полнодисковое шифрование

Прежде чем начать рассмотрение по существу, уделим внимание терминологии. Полнодисковое шифрование — наиболее устоявшийся в данной предметной области термин, обозначающий процесс, который реализуется модулями защиты, функционирующими между драйвером диска и файловой системой, основывается на применении криптографических методов и предназначается для обеспечения безопасности данных, хранящихся на жёстких дисках. Первая часть термина («полно») объясняется тем, что данный процесс предполагает защиту всего пространства диска, выделенного для хранения полезной нагрузки (прикладных данных). При этом вторая и третья части термина («диск» и «шифрование») могут вызвать вопросы. Во-первых, из-за того, что на сегодняшний день в основе устройств, применяемых для хранения информации, не обязательно лежит набор физических дисков. Во-вторых, из-за того, что защита хранимых данных, как показано далее, не всегда ограничивается только шифрованием. Однако в рамках настоящей работы будем использовать именно этот термин, поскольку он является наиболее близким русскоязычным аналогом устоявшегося в зарубежной литературе термина «Full Disk Encryption» [7, 4]. Далее для краткости будем использовать соответствующую аббревиатуру FDE.

### 2.1. Основные понятия

Модуль FDE перехватывает запрос от файловой системы, преобразует переданные в его составе данные и осуществляет взаимодействие с драйвером диска для реализации функций безопасности и собственно записи данных на диск. Для реализации «прозрачной» защиты данных интерфейс FDE должен полностью совпадать с интерфейсом драйвера жёсткого диска той ОС, в которой он функционирует. Таким образом, файловая система «не знает» о том, что данные на используемом ею жёстком диске защищены. Схема взаимодействия файловой системы с драйвером жёсткого диска при наличии модуля FDE представлена на рис. 2.

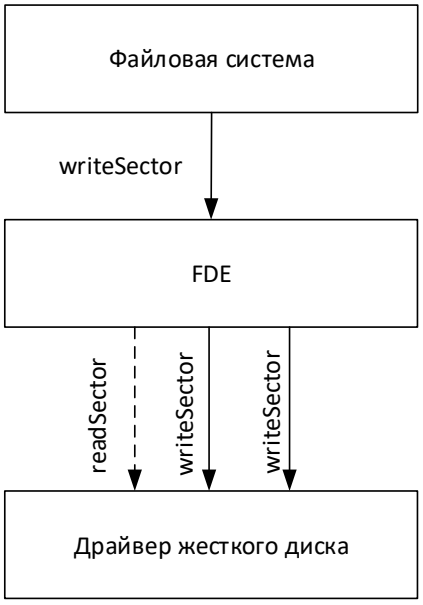



Рис. 2. Взаимодействие файловой системы с драйвером жесткого диска при наличии модуля FDE

При обработке запроса от файловой системы модуль FDE может посылать к драйверу диска уже несколько запросов на чтение и запись. Например, запрос на запись одного сектора, сделанный файловой системой, может потребовать от модуля FDE одной операции чтения для получения текущего значения количества осуществлённых операций записи данного сектора (для контроля нагрузки на ключ или формирования уникального инициализирующего вектора) и двух операций записи (одна для записи непосредственных данных сектора, вторая — для записи нового значения имитовставки сектора). Отдельно отметим, что FDE-модуль должен допускать параллельное осуществление нескольких операций с прикладными данными диска.

**Замечание 1.** Использование FDE-модулем для хранения служебной информации исключительно пространства защищаемого диска, посекторный доступ к которому реализуется некоторым драйвером, является типичным сценарием, который наиболее удобен для объяснения основных концепций. Однако в общем случае необходимо лишь, чтобы FDE-модуль предоставлял посекторный доступ компонентам более верхнего уровня. То, как FDE-модуль реализует такую «посекторную» абстракцию, может зависеть от особенностей решаемой задачи. Например, FDE-модуль может сохранять данные в обычный файл, доступ к которому может быть реализован с помощью штат-

ных возможностей файловой системы. При этом для хранения служебных данных FDE-модуль может использовать некоторую базу данных, порядок взаимодействия с которой может существенно отличаться от посекторного. В таких случаях определения некоторых эксплуатационных характеристик, обсуждаемых в п. 2.2, должны быть скорректированы. Далее все рассуждения проводятся для типичного случая работы FDE-модуля со стандартным драйвером диска, предоставляющим посекторный доступ.

В настоящей работе рассмотрим только те FDE-схемы, для которых как минимум часть ключевого материала, используемого для обеспечения целевых функций безопасности, хранится на защищённом носителе (токене). Альтернативой являются схемы, стойкость которых основана на запоминаемом пользователем пароле. К сожалению, такие схемы не позволяют достичь интересующего авторов уровня стойкости. Необходимый FDE-модулю размер памяти токена, доступной для чтения и, возможно, записи, является важной характеристикой FDE-модуля.

Для большинства (контрпримеры авторам не известны) современных FDE-схем можно выделить четыре этапа работы. Они перечислены ниже. Будем рассматривать только те схемы, для которых справедливо такое разделение.

1) *Этап инициализации* (*Setup*):
   на данном этапе FDE-модулю на вход подаётся диск в исходном незащищённом виде; модуль осуществляет все действия, необходимые для его последующей защиты, а именно его разметку с выделением, возможно, области для своих служебных данных, генерацию ключей, начальное зашифрование данных, сохранение необходимой информации на токене и т. п.
2) *Этап начала сеанса* (*Mount*):
   предоставление FDE-модулю возможности взаимодействия с токеном, содержащим ключевой материал, с помощью которого защищён диск; разворачивание в оперативной памяти ОС ключевой системы, необходимой для работы с FDE-модулем; чтение с токена дополнительных данных; осуществление других операций, необходимых для обеспечения целевых свойств безопасности и перехода к следующему этапу.
3) *Основной этап* (*Session*):
   взаимодействие файловой системы с драйвером жёсткого диска (чтение и запись данных) через модуль FDE.
4) *Этап завершения сеанса* (*Unmount*):
   осуществление операций, необходимых для обеспечения целевых свойств безопасности и завершения работы с FDE-модулем; завершение работы с токеном (например, загрузка модифицированных за время сеанса дополнительных данных и отсоединение токена от рабочей машины).

### 2.2. Эксплуатационные свойства

Перечислим эксплуатационные характеристики и особенности, которые необходимо учитывать при проектировании и анализе FDE-схем. Сначала отметим одну особенность функционирования FDE-модулей. Учитывая высокие требования к эффективности, при реализации FDE-модулей могут использоваться различные приёмы, позволяющие увеличить скорость их работы. Так, некоторые данные (например, данные, позволяющие проверять целостность секторов) могут заранее загружаться с диска в оперативную память. Таким образом, при определении характеристик FDE-схем

нужно учитывать возможную зависимость их эффективности от размера доступной оперативной и, возможно, защищённой памяти.

При проектировании FDE-схем должны учитываться эксплуатационные требования как к основному этапу их работы (сеансу), так и к другим трём этапам. Так, сложно представить ситуацию, допускающую монтирование диска в течение 30 мин.

**Используемая память.** Безусловно, характеристиками любой FDE-схемы являются необходимые ей для работы размер защищённой памяти, размер оперативной памяти и относительный размер её служебных данных (т. е. отношение размера служебных данных к размеру защищаемого полезного пространства диска). Учитывая то, что было сказано о возможностях оптимизации работы FDE-схем, при характеризации конкретной схемы необходимо использовать некоторую относительную величину, например, может быть приведено отношение размера используемой оперативной памяти к размеру области диска, работа с которым может быть оптимизирована.

**Операционная трудоёмкость.** Под этим термином понимается количество операций чтения и записи секторов, осуществляемых драйвером диска, при реализации FDE-модулем своих функций на разных этапах работы. Эти операции могут существенно отличаться по быстродействию в зависимости от используемой аппаратной платформы. Они выделяются в отдельную категорию ещё и потому, что зачастую ощутимо влияют на износ диска, а их допустимое количество заметно ограничено.

**Вычислительная трудоёмкость.** Данная характеристика является относительно стандартной и подразумевает объём вычислений, производимых FDE-модулем для осуществления операций на разных этапах работы.

**Устойчивость к сбоям.** Для устройств хранения данных крайне важным является вопрос надёжности: аварийные ситуации различного характера должны минимально влиять на хранимые данные. Использование FDE-модуля может различным образом влиять на это. Например, если при внезапном отключении питания актуальные ключи шифрования данных хранились только в оперативной памяти, то все данные диска могут стать недоступными. При этом необходимо учитывать не только получение самих данных, но и сохранение доверия к ним (в случае обеспечения целостности). Опишем некоторые практически актуальные аварийные ситуации, которые могут возникнуть во время работы FDE-модуля.

— *Отключение питания.* FDE-модуль мгновенно прекращает свою работу, используемая им информация в оперативной памяти утрачивается.
— *Отключение диска.* FDE-модуль теряет возможность осуществлять операции чтения и записи как минимум части данных, хранящихся в долговременной памяти. В качестве примера можно привести гипотетический FDE-модуль, оперирующий с данными, которые хранятся на нескольких дисках, один из которых отсоединяется пользователем.
— *Отключение защищённой памяти.* Например, пользователь может по ошибке отсоединить токен от вычислительной машины, на которой в этот момент работает FDE-модуль. В этом случае FDE-модуль теряет возможность оперировать с защищённой памятью.

Дополнительно отметим аварийную ситуацию, которая может реализоваться в любой момент существования диска.

— *Ошибки хранения данных на физическом уровне.* Например, из-за износа диска некоторый сектор может перестать быть доступным для чтения. Это может привести как к невозможности расшифровать часть данных, так и к потере доверия

к их целостности. Довольно простой мерой предотвращения таких ситуаций может быть дублирование данных диска (данные хранятся на двух дисках одинакового размера, операции дублируются).

Некоторые FDE-модули могут предусматривать особые процедуры, связанные с аварийными ситуациями. Это приводит к выделению новых этапов работы FDE-модулей, примеры которых приведены далее:

5) *Этап аварийной работы* (*Emergency*):
начинается в том момент, когда реализуется аварийный сценарий, допускающий продолжение работы в принципе. FDE-модуль может как продолжить осуществлять прикладные запросы (если это осмысленно), так и выполнять операции, необходимые для максимально корректного и безопасного завершения работы.
6) *Этап восстановления* (*Recovery*):
осуществляются операции, необходимые для восстановления работы с защищаемыми данными после сбоя (этапы могут различаться в зависимости от типа произошедшего сбоя).

### 2.3. Криптографические свойства

Говорить о криптографических свойствах имеет смысл только в рамках набора условий и предположений, определяющих возможности и цели потенциальных противников и обычно объединяемых под названием «модель противника». Подробно это понятие и один из подходов к его формализации рассмотрены в [9, 10]. Модель противника состоит из трёх компонент: типа атаки, модели угрозы и доступных противнику информационных и вычислительных ресурсов. Третья компонента обычно обсуждается уже на этапе анализа конкретных криптосистем, а не в рамках работ о предметной области в целом, но одну её особенность для систем полнодискового шифрования необходимо отметить.

Под информационными ресурсами противника обычно понимается количество данных, которые он потенциально может получить в процессе работы криптосистемы (шифртексты, имитовставки, электронные подписи и т. п.). Пути получения этой информации и её характер могут быть очень разными: перехват зашифрованных сообщений в канале связи, получение информации о внутренних состояниях криптосистемы по побочным каналам и т. д. Одним из самых ярких наглядных примеров того, что получение противником большого количества такой информации может приводить к серьёзным уязвимостям, является атака Sweet32 [11] на протокол TLS. Общим методом исключения таких возможностей противника является ограничение так называемой «нагрузки на ключ», т. е. количества данных, которые могут быть обработаны на одном ключе (этот приём и само понятие «нагрузка на ключ» рассматривается в [12, 13]; неформальное рассмотрение можно найти в [14, 15]). При синтезе и анализе FDE-модулей данному аспекту необходимо уделять особое внимание, так как период интенсивного функционирования и, как следствие, использования ключей такими криптосистемами может исчисляться годами и порой ограничивается лишь сроками службы устройств, используемых для хранения сопутствующих данных.

Две другие части модели противника определяют возможности противника по взаимодействию с криптосистемой (*тип атаки*) и его цели по нарушению целевых для неё свойств безопасности (*модель угрозы*). Рассмотрим каждую из этих компонент для систем полнодискового шифрования.

Тип атаки

В рамках функционирования FDE-модуля выделяются следующие компоненты, связанные с хранением и обработкой критичной для криптосистемы информации:

— защищённая память;
— оперативная память устройства, на котором функционирует FDE-модуль (о ней есть смысл говорить только во время работы FDE-модуля);
— совокупность долговременно хранимых данных, используемых FDE-модулем для реализации абстракции защищаемого диска (далее для краткости будем называть эти данные просто защищаемым диском).

Потенциально противник может взаимодействовать со всеми перечисленными компонентами, например получать какую-либо информацию по побочным каналам. Однако в рамках настоящей работы подробно рассмотрим взаимодействие противника только с защищаемым диском.

Особо отметим, что в настоящей работе не рассматриваются вопросы целостности и безопасной загрузки исполняемого кода FDE-модуля и сопутствующего программного обеспечения. Так, предполагается, что загрузка операционной системы осуществлена под защитой, например, некоторого аппаратно-программного модуля доверенной загрузки.

**Взаимодействие с защищаемым диском.** Противник может взаимодействовать с защищаемым диском, читая и записывая данные. При обсуждении этих возможностей будем следовать традиционному для криптографии подходу, состоящему в максимизации этих возможностей (например, возможности писать любые данные в любые секторы, а не только данные какого-либо специального содержания в какие-либо конкретные секторы). Это связано с тем, что спектр возможных условий эксплуатации систем полнодискового шифрования, как и любых криптосистем более или менее общего характера, крайне широк и может быть фиксирован весьма частично.

Для операций чтения и записи можно выделить две характеристики: уровень осуществления операции и период осуществления операции.

Под уровнем осуществления операции будем понимать, с помощью какого модуля операционной системы противник осуществляет конкретную операцию. Выделяются две основные возможности.

— *Операции уровня FDE-интерфейса.* Противник может сделать запрос к FDE-модулю на чтение или запись данных защищаемого диска (любых данных в любые секторы диска). На практике доступ такого типа к диску противник может получить, например, навязав легальному пользователю сохранение некоторого файла на диске вместо другого файла, расположение которого на этом диске противнику известно.
— *Операции уровня Raw-интерфейса.* Противник осуществляет операции с помощью драйвера диска, которым пользуется и сам FDE-модуль для сохранения на диск всех необходимых данных. Так, с помощью этого интерфейса противник может прочитать или переписать зашифрованные секторы или области служебных данных FDE-модуля. Наиболее распространённым в литературе по данной тематике является практический пример с кражей защищаемого диска — в данном случае противник может считывать и записывать информацию диска, взаимодействуя с ним как с незащищённым.

Заметим, что FDE-интерфейс, в отличие от Raw-интерфейса, не позволяет противнику получить прямой доступ к служебным данным FDE-модуля.

В результате доступа к обозначенным интерфейсам противник может навязывать открытые данные с помощью FDE-интерфейса и считывать через Raw-интерфейс информацию о полученных шифртекстах и имитовставках, а также записывать на диск подобранные шифртексты и, например, имитовставки с помощью Raw-интерфейса и получать соответствующий результат расшифрования через FDE-интерфейс. Это является практическим отражением возможностей противника, предоставляемых ему формальными моделями типа CPA (Chosen Plaintext Attack) и CCA (Chosen Ciphertext Attack) [9, 16].

Период осуществления операции — это то время, когда у противника есть возможность осуществить операцию. Так, делать операции с помощью FDE-интерфейса противник может только на этапах Session и, возможно, Emergency.

В случае Raw-интерфейса обычно предполагается, что противник может осуществлять Raw-чтение в любой момент времени и Raw-запись данных, когда FDE-модуль не работает. При этом с возможностью Raw-записи во время работы FDE-модуля ситуация не так однозначна.

Это объясняется тем, что такая возможность может существенно усложнить задачу обеспечения целостности хранимых данных. Например, если у противника этой возможности нет, то для дисков малого объёма целостность может быть проверена на этапе Mount, что позволит исключить часть операций на этапе Session, от которого обычно требуется максимальная производительность. Если противник может записывать через Raw-интерфейс на этапе Session, то предвычисления на этапе Mount теряют смысл, так как целостность может быть нарушена в любой момент после этапа Mount. Стоит заметить, что на практике такая возможность противника является довольно сильной, а более реалистичны сценарии, когда противник получает полный доступ к диску не во время работы с ним легитимного пользователя, а между сеансами, например в результате обхода парольной системы аутентификации операционной системы.

**Атаки, связанные со сбоями.** Специфическим аспектом функционирования FDE-модулей являются различные сбои, некоторые из которых рассмотрены в п. 2.2. При этом нельзя исключать, что противник может использовать сбои для нарушения целевых свойств безопасности системы, поэтому это необходимо учитывать в рамках криптографического анализа FDE-модулей. Заметим, что при формировании модели противника для FDE-модуля могут учитываться и задействоваться следующие приёмы и организационные меры, которые позволяют ограничить возможности противника по провоцированию сбоев:

— особые требования к хранению диска после отключения питания до начала этапа Recovery для исключения доступа к нему противника;
— требование использования источников бесперебойного питания при работе FDE-модулей;
— использование счётчика числа сбоев и требование смены ключевого материала и нового прохождения этапа Setup по достижении им порогового значения.

Модель угрозы

Целевыми свойствами безопасности, для обеспечения которых предназначен FDE-модуль как криптографический механизм, являются конфиденциальность и целостность хранимых данных. Хотя данные свойства в криптографии стандартны и достаточно глубоко исследованы, специфика задачи защиты хранимых данных приводит к появлению у них новых аспектов и особенностей.

Прежде чем переходить к подробному рассмотрению каждого из указанных свойств, необходимо более явно обозначить объект, для которого эти свойства необходимо обеспечивать. Неформально, под защищаемыми данными подразумеваются все прикладные данные, которые записываются на диск через FDE-модуль с момента его инициализации. Можно выделить два подхода к представлению совокупности этих данных в виде потоков сообщений:

1) Представление в виде одного потока сообщений, где одно сообщение — это совокупность всех прикладных данных, записанных на диск через FDE-модуль после очередной операции записи сектора, инициированной файловой системой. Одно сообщение удобно представлять в виде упорядоченного набора, где элементом с номером $i$ являются прикладные данные, записанные в $i$-й сектор. Такой набор также можно сравнить с «моментальным снимком», где сохранилось состояние данных диска, которое было неизменным в течение некоторого промежутка времени, т. е. каждый следующий «снимок» отличается от предыдущего ровно в одном секторе. Принцип данного подхода представлен на рис. 3.

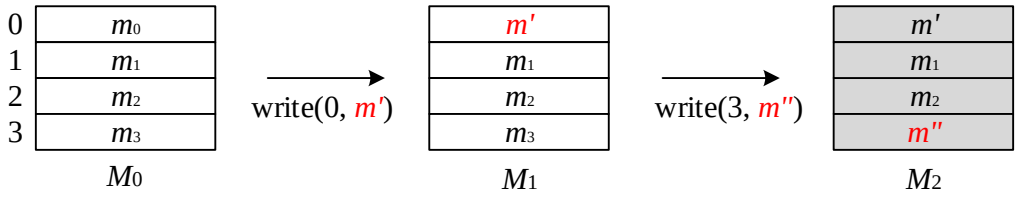



Рис. 3. Представление защищаемых данных в виде одного потока сообщений

2) Представление в виде фиксированного количества независимых потоков сообщений, каждый из которых соответствует конкретному прикладному сектору. В этом случае одним сообщением в конкретном потоке являются прикладные данные, которые записываются в соответствующий сектор, а каждое новое сообщение в этом потоке возникает после очередной операции записи именно в этот сектор. Данный подход проиллюстрирован на рис. 4.

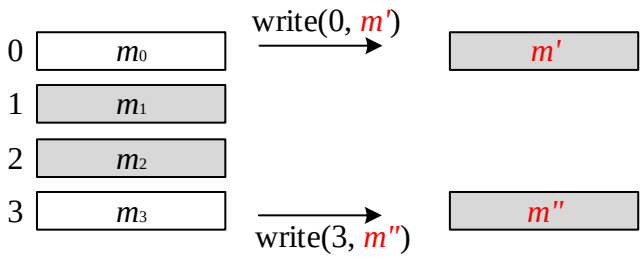



Рис. 4. Представление защищаемых данных в виде набора потоков сообщений

Далее, говоря про конфиденциальность или целостность, будем иметь в виду конфиденциальность или целостность одного потока сообщений, каждое из которых является либо совокупностью всех данных диска, либо данными конкретного сектора. При

этом в зависимости от того, что именно является сообщением, то или иное свойство безопасности может иметь различный практический смысл. Поэтому для различения градаций какого-либо свойства будем говорить, что это свойство обеспечивается *на уровне диска*, если сообщением является совокупность всех прикладных данных, или *на уровне сектора*, если сообщением являются прикладные данные конкретного сектора. Рассмотрим подробнее каждое из целевых свойств безопасности.

**Конфиденциальность.** Это является основным свойством безопасности хранимых на диске прикладных данных, которое должен обеспечивать FDE-модуль. Отметим особенность обеспечения его в контексте полнодискового шифрования. На практике на защищаемом диске могут храниться как важные документы, содержимое которых должно оставаться в секрете многие годы, так и открытые данные или данные, которые становятся публичными по прошествии небольшого промежутка времени. Таким образом, в одном и том же секторе диска в разные моменты времени, а также одновременно в разных секторах диска может быть записана информация, к которой предъявляются различные требования по обеспечению конфиденциальности.

Под конфиденциальностью хранимых на диске прикладных данных, представленных в виде одного или нескольких потоков сообщений, будем неформально понимать невозможность раскрытия противником какой-либо информации об этих сообщениях. Ориентиром при формализации этого свойства могут служить модели на основе неотличимости (например, IND-CPA [16]). В зависимости от выбранного представления диска (т. е. от того, что именно понимается под сообщением) некоторая информация о прикладных данных может заведомо раскрываться противнику. Выделим следующие степени конфиденциальности хранимых данных по характеру доступной противнику информации:

— Конфиденциальность на уровне диска. В этом случае противник не получает никакой информации о хранимых на диске прикладных данных. Уточним, что, помимо неполучения информации о самих хранимых в секторах данных, противник также не должен узнавать номер изменяемого сектора. Обеспечение такого свойства представляется крайне трудозатратным — одним из подходов является перешифрование всего диска после операции записи любого сектора. Это неэффективно в случае дисков большого размера и не позволяет параллельно обрабатывать операции с разными секторами.
— Конфиденциальность на уровне сектора. В этом случае противник не получает никакой информации о прикладных данных, хранимых в конкретном секторе. Отметим, что обеспечение конфиденциальности на уровне сектора для всех прикладных секторов диска неэквивалентно обеспечению конфиденциальности на уровне диска. В сравнении с предыдущим свойством у противника появляется информация о том, в какие именно секторы осуществляется обращение. Несмотря на то, что схемы, обеспечивающие конфиденциальность на уровне сектора, являются более эффективными, такой конфиденциальности не всегда достаточно. Так, например, если на защищаемом диске хранится карта военных действий, а каждый сектор соответствует определённому пункту местности, раскрытие информации о том, с какими секторами идёт работа, может быть критичным.

Необходимым условием обеспечения полноценной конфиденциальности является увеличение размера шифртекста по сравнению с размером открытого текста (например, за счёт вектора инициализации, см. подробнее [16]), что возможно только при наличии дополнительной памяти для хранения служебных данных.

Если использование дополнительной памяти невозможно, наилучшей степенью конфиденциальности, которую удастся обеспечить, является так называемая «конфиденциальность по модулю повторений» [17]. В этом случае противник дополнительно детектирует факт совпадения сообщений в потоке. Иногда раскрытие такой информации может быть критичным. Предположим, например, что каждый день в определённый сектор на диске записывается актуальный на следующий день курс акций компании. Эта информация публикуется на следующий день, а до этого момента должна оставаться в секрете. Если противник устанавливает факт совпадения актуальных данных сектора с данными, записанными некоторое время назад, то он немедленно получает информацию о курсе акций на завтрашний день.

В предыдущем разделе отмечено, что противник имеет возможность модифицировать данные на диске. Это означает, что помимо традиционных атак (навязывание противником открытого текста), при анализе свойств безопасности FDE-модуля должны учитываться атаки, опирающиеся на возможность противника записывать на диск специальным образом подобранные шифртексты. Рассмотрим несложный пример такой атаки.

Предположим, что в некоторый момент времени в определённый сектор была записана критически важная информация, конфиденциальность которой необходимо обеспечивать в течение большого промежутка времени. Предположим также, что далее в этот же сектор была записана менее важная информация, результат расшифрования которой противнику может быть доступен; например, это может быть текст отчёта, который становится известным в короткие сроки. Допустим, что через некоторое время после записи менее секретных данных у противника появляется возможность недетектируемо заменить соответствующий им шифртекст на корректный шифртекст, соответствующий критически важной информации. Тогда при опубликовании отчёта критически важная информация будет раскрыта.

Опишем один из подходов к осмыслению потенциальных возможностей противника в рассмотренной атаке. Введём градации свойства конфиденциальности по времени, в течение которого неактуальные данные являются уязвимыми. Под актуальными данными будем понимать последнее сообщение в потоке, а моментом неактуальности данных будем называть время появления следующего за ними сообщения в потоке. Для конкретных данных (сообщения в потоке) определим свойство *конфиденциальности по модулю* $\delta$ следующим образом. В течение промежутка времени $\delta$ с момента неактуальности данные являются уязвимыми и нарушение их конфиденциальности не считается угрозой, в остальное время их конфиденциальность обеспечивается. Рассмотрим некоторые частные случаи:

— $\delta = 0$. Конфиденциальность данных, записанных на диск в разные моменты времени, обеспечивается в течение всего времени работы FDE-модуля, т. е. угрозой считается получение противником информации о произвольном сообщении в потоке независимо от того, как давно оно было записано;
— $\delta > 0$. Данные являются уязвимыми в течение промежутка времени $\delta$ с момента неактуальности. При этом получение противником информации об актуальных данных, а также о «старых» неактуальных данных считается угрозой. Отметим, что при $\delta = \infty$ угрозой является лишь получение противником информации об актуальных данных.

Таким образом, чем меньше значение $\delta$, тем меньше промежуток времени, в течение которого данные являются уязвимыми, а потому тем сильнее свойство конфиденциальности по модулю $\delta$.

Вернёмся к атаке. Если противник смог недетектируемо заменить содержимое сектора спустя промежуток времени $t$ с момента неактуальности секретных данных, то он нарушил свойство конфиденциальности по модулю $\delta$ для всех $\delta \leqslant t$ (в частности, конфиденциальности по модулю 0). Однако конфиденциальность по модулю $\infty$ в этом случае не нарушена, так как по построению атаки противник не получает информации об актуальных данных соответствующего сектора.

Мы вводим свойство конфиденциальности по модулю $\delta$ таким образом, что его нарушение становится возможным, только если противник осуществляет атаки, подразумевающие модификацию данных на диске. Поэтому обеспечение конфиденциальности по модулю $\delta$ свидетельствует о том, что противник имеет возможность недетектируемо модифицировать данные на диске только в течение промежутка времени $\delta$ с момента их неактуальности.

Если схема полнодискового шифрования обеспечивает конфиденциальность по модулю $\delta > 0$, достичь конфиденциальности по модулю 0 можно с помощью организационных мер. Например, можно потребовать, чтобы после каждого выключения FDE-модуля пользователь контролировал диск в течение промежутка времени $\delta$, не допуская тем самым возможности противника модифицировать данные на нём.

**Целостность.** Под целостностью данных понимается невозможность внесения противником недетектируемых изменений в данные, записанные на диск. Важность обеспечения данного свойства часто недооценивают, что может быть связано со следующей особенностью задачи защиты дисков. В отличие от задачи защиты канала связи между двумя пользователями, в данном случае «отправляющая» (записывающая) и «принимающая» (считывающая) стороны являются одним и тем же объектом — файловой системой, за которой чаще всего стоит человек. Поэтому потенциально человек способен обнаружить изменения в считываемых данных, так как сам когда-то записывал их на диск через FDE-модуль. Однако в реальности на диске могут храниться как очень большие объёмы данных, следить за неизменностью которых человек просто не способен, так и служебная информация приложений, которая человеку как пользователю неизвестна. Ярким примером серьёзной уязвимости, которая может появиться при отсутствии целостности, является возможность незаметно внести изменения в хранящийся на диске исходный код некоторой программы, которые не повлияют на её компилируемость, но приведут к критичному изменению функционирования.

Принципиальным моментом в обеспечении целостности хранимых данных является обеспечение их «актуальности», т. е. легитимный пользователь должен быть уверен, что при чтении он получит те же данные, которые были сохранены на диск в результате предыдущей операции записи. С точки зрения представления данных в виде потока сообщений целостность означает целостность именно последнего (актуального) сообщения в потоке в каждый момент времени. Проводя аналогию с каналами связи, ситуация, когда данные сектора перестают быть актуальными, соответствует ситуации, когда сообщение либо так и не дошло до получателя, либо уже им прочитано. Соответственно изменение этого сообщения противником не имеет смысла.

Как и для конфиденциальности, в зависимости от точки зрения на защищаемые данные смысл целостности может отличаться:

1) Целостность на уровне диска. Под данным свойством понимается невозможность внесения каких-либо изменений в данные диска как упорядоченного набора, в том числе перемещение данных из одного сектора в другой.
2) Целостность на уровне сектора. Под данным свойством подразумевается невозможность внесения каких-либо изменений в данные конкретного сектора (в том числе невозможность замены данных этого сектора на данные других секторов). Отметим, что в отличие от конфиденциальности обеспечение целостности на уровне сектора для каждого сектора эквивалентно обеспечению целостности на уровне диска.

Обеспечение целостности невозможно без увеличения длины шифртекста относительно длины открытого текста (подробнее см. [7]), что влечёт необходимость наличия места для хранения служебных дополнительных данных, например имитовставок, и как следствие — к усложнению FDE-модулей и замедлению скорости обработки данных. Поэтому некоторые FDE-модули обеспечивают так называемую «псевдоцелостность», не требующую выделения дополнительного места на диске. Под псевдоцелостностью понимается невозможность внесения контролируемых изменений в сообщение, т. е. любые изменения шифртекста приведут к тому, что расшифрованный текст будет выглядеть как случайное сообщение. Псевдоцелостность на уровне диска отличается от псевдоцелостности на уровне сектора: в первом случае изменение шифртекста любого сектора должно приводить к непредсказуемым изменениям всех прикладных данных, а во втором — к непредсказуемым изменениям только прикладных данных, хранящихся в соответствующем изменённом секторе. Хотя данное свойство не является целостностью в стандартном смысле, его обеспечение также можно отнести к защите от недетектируемого изменения данных. Действительно, чем к большим изменениям прикладных данных приводит изменение шифртекста, тем больше шансов, что такое изменение приведёт к ошибке на прикладном уровне и, таким образом, не останется незамеченным.

Для целостности, как и для конфиденциальности, можно ввести градацию по степени обеспечения актуальности данных с помощью понятия *целостности по модулю* $\delta$. Неформально данное свойство означает, что противник может недетектируемо заменить актуальные данные только на те данные, которые хранились на диске не более чем $\delta$ единиц времени назад. С точки зрения потока сообщений оно означает возможность противника заменить последнее в потоке сообщение на любое из нескольких предыдущих, и только на него. Частные случаи:

— $\delta = 0$. Гарантируется, что при чтении пользователь получает доступ к тем же данным, которые были записаны им при последних обращениях к секторам, и угрозой считается внесение любых изменений в актуальное сообщение. Обеспечение целостности по модулю 0 на уровне сектора для всех прикладных секторов эквивалентно обеспечению целостности по модулю 0 на уровне диска. Необходимым эксплуатационным условием для обеспечения данного свойства является наличие защищённой памяти, модифицировать которую противник не может, так как данное свойство связано с необходимостью передачи некоторого внутреннего состояния от сессии к сессии [7];
— $\delta > 0$. Угрозой считается замена «актуальных» сообщений на достаточно «старые» сообщения потока или на новые сообщения, которых в потоке нет. При $\delta > 0$ обеспечение целостности по модулю $\delta$ на уровне диска неэквивалентно обеспечению целостности по модулю $\delta$ на уровне сектора для всех прикладных секторов. Дей-

ствительно, при обеспечении целостности по модулю $\delta$ на уровне диска у противника появляется лишь возможность недетектируемо заменять данные всех секторов на диске на данные, которые одновременно хранились в этих секторах диска в некоторый предыдущий момент времени. При обеспечении целостности по модулю $\delta$ на уровне сектора у противника остаётся возможность «откатывать» данные различных секторов к неактуальному состоянию независимо друг от друга, т. е. противник может привести диск в новое состояние, комбинируя данные, которые хранились в различных секторах в разные моменты времени. При $\delta = \infty$ актуальность данных не обеспечивается и угрозой считается только недетектируемая замена актуального сообщения на сообщения, которые никогда не встречались в потоке (не хранились в секторе или на диске).

Необходимо отметить, что хотя целостность по модулю $\delta$ сама по себе может быть важным свойством безопасности, её наличие позволяет также обеспечить конфиденциальность в условиях, когда противник может изменять данные на диске. Действительно, для обеспечения конфиденциальности по модулю $\delta$ достаточно обеспечить конфиденциальность по модулю $\infty$ и целостность по модулю $\gamma$, где $\gamma \leqslant \delta$.

### 2.4. Некоторые типичные сценарии использования

Нельзя исключать, что добиться приемлемых значений всех перечисленных характеристик для всех возможных сценариев использования защищённого диска невозможно. Решением проблемы является создание нескольких FDE-схем, каждая из которых оптимизирована для конкретных более узких условий эксплуатации. Далее мы, не претендуя на полноту, описываем некоторые типичные на сегодняшний день сценарии использования защищённых дисков и их особенности, которые могут облегчить процесс принятия синтезных решений.

**Персональный съёмный диск.** Примером такого диска является USB-флэш-диск, используемый для хранения конфиденциальной информации личного и/или рабочего характера. Особенности работы с диском такого типа:

— относительно небольшой объём защищаемых данных (до 256 Гбайт);
— небольшой объём оперативной памяти машин, на которых осуществляется работа с диском (например, единицы гигабайт);
— относительно непродолжительные сеансы работы и большое количество операций начала и завершения сессии; например, пользователь может подключать диск к машине в начале рабочего дня и отключать в конце;
— небольшой совокупный размер данных, с которыми производятся операции в течение сеанса или, по крайней мере, продолжительного промежутка времени в рамках сеанса. Так, обычно пользователь оперирует только с данными, относящимися к актуальному проекту, а данные завершённых проектов не задействуются;
— ненулевая вероятность сбоев всех типов. Так, в случае частого использования возможно отключение пользователем диска от рабочей машины без процедуры завершения сеанса.

**Хранилища центров обработки данных.** Этот сценарий предполагает защищённое хранение огромного объёма данных, доступ к которым серверная операционная система одновременно предоставляет большому числу клиентов. Особенности работы с диском в рамках такого сценария:

— большой объём защищаемых данных (от 2 Тбайт);
— большой объём оперативной памяти сервера, осуществляющего работу с диском;

— малое число операций начала и завершения сеанса работы и крайне продолжительные сессии;
— большой объём данных, с которыми производятся операции в течение даже не самого продолжительного периода времени; при этом расположение запрашиваемых секторов имеет преимущественно случайный характер;
— минимальная вероятность сбоев. Дублирование хранения обеспечивает защиту от ошибок на физическом уровне и от внезапных потерь доступа к диску; использование систем бесперебойного питания минимизирует риск внезапного отключения питания машины.

## Заключение

В работе описаны и проанализированы основные эксплуатационные и криптографические свойства FDE-схем защищённого хранения данных в современных операционных системах, а также особенности некоторых типичных сценариев использования защищённых с помощью FDE-схем дисков. Эти сведения могут быть использованы при проектировании FDE-схем.

При этом стоит отметить, что приведённые рассуждения о криптографических свойствах таких систем нацелены на их неформальное понимание и могут служить лишь ориентиром при формальном задании модели противника (например, на основе подхода, описанного в [9]) и последующей оценке стойкости таких схем.

Проведённый анализ вытекающих из практики требований к режимам полнодискового шифрования является необходимым для формирования научно-технической базы синтеза таких режимов. Начатые в настоящей работе исследования продолжатся в направлении анализа как существующих режимов, так и предлагаемых к стандартизации в Техническом комитете «Криптографическая защита информации» (ТК 26).

Авторы выражают глубокую благодарность своим коллегам Л. О. Никифоровой, А. А. Русеву, Е. С. Грибоедовой, Л. А. Сониной и Д. А. Щербакову за плодотворные обсуждения, ценные замечания и конструктивную критику.

## ЛИТЕРАТУРА

1. Рекомендации по стандартизации Р 1323565.1.012-2017 «Информационная технология. Криптографическая защита информации. Принципы разработки и модернизации шифровальных (криптографических) средств защиты информации». М.: Стандартинформ, 2017.
2. Об утверждении Требований к средствам электронной подписи и Требований к средствам удостоверяющего центра. Приказ ФСБ РФ от 27 декабря 2011 г. № 796.
3. *Зима В. М., Клюев А. В., Литвинов О. А. и др.* Основы защиты информации от несанкционированного доступа в автоматизированных системах конфиденциального делопроизводства // Тр. СПИИРАН. 2006. Вып. 3. Т. 2. С. 84–95.
4. *Khati L.* Full Disk Encryption and Beyond. Diss. Cryptography and Security [cs.CR]. Universite PSL; ENS Paris — Ecole Normale Superieure de Paris, 2019. 182 p.
5. *Broz M.* Authenticated and Resilient Disk Encryption. PhD thesis. Brno: Masaryk University, 2018.
6. *Damgard I. and Dupont K.* Universally Composable Disk Encryption Schemes. IACR Cryptology ePrint Archive. 2005. `https://eprint.iacr.org/2005/333.pdf`.
7. *Gjosteen K.* Security notions for disk encryption // LNCS. 2005. V. 3679. P. 455–474.
8. `https://integralmemory.com`.

9. *Алексеев Е. К., Ахметзянова Л. Р., Зубков А. М. и др.* Об одном подходе к формализации задач криптографического анализа // Матем. вопр. криптогр. 2020 (в печати).
10. *Алексеев Е. К., Ахметзянова Л. Р., Карпунин Г. А. и др.* Что плохого можно сделать, неправильно используя криптоалгоритмы? Доклад на лектории симпозиума CTCrypt'2019. https://ctcrypt.ru/files/files/2019/materials/29_Alekseyev.pdf.
11. *Bhargavan K. and Leurent G.* On the practical (in-)security of 64-bit block ciphers. Collision attacks on HTTP over TLS and OpenVPN // Proc. CCS'16, October 24–28, 2016, Vienna, Austria. P. 456–467. https://sweet32.info/SWEET32_CCS16.pdf.
12. *Smyshlyaev S.* Re-keying Mechanisms for Symmetric Keys. RFC 8645. August 2019. https://tools.ietf.org/html/rfc8645.
13. *Akhmetzyanova L. R., Alekseev E. K., Oshkin I. B., and Smyshlyaev S. V.* Increasing the Lifetime of Symmetric Keys for the GCM Mode by Internal Re-keying. Cryptology ePrint Archive: Report 2017/697.
14. *Алексеев Е. К., Ахметзянова Л. Р., Мешков Д. А. и др.* О нагрузке на ключ. Ч. 1. Блог ООО «КРИПТО-ПРО». 2017. http://cryptopro.ru/blog/2017/05/17/o-nagruzke-na-klyuch-chast-1.
15. *Алексеев Е. К., Ахметзянова Л. Р., Мешков Д. А. и др.* О нагрузке на ключ. Ч. 2. Блог ООО «КРИПТО-ПРО». 2017. http://cryptopro.ru/blog/2017/05/29/o-nagruzke-na-klyuch-chast-2.
16. *Bellare M. and Rogaway P.* Introduction to Modern Cryptography. 2005. https://web.cs.ucdavis.edu/~rogaway/classes/227/spring05/book/main.pdf.
17. *Bellare M., Boldyreva A., and O'Neill A.* Deterministic and efficiently searchable encryption // LNCS. 2007. V. 4622. P. 535–552.

## REFERENCES

1. Informacionnaya tekhnologiya. Kriptograficheskaya zashchita informacii. Principy razrabotki i modernizacii shifroval'nyh (kriptograficheskih) sredstv zashchity informacii [Information Technology. Cryptographic Data Security. Principles of Creation and Modernization for Cryptographic Modules. Recommendations for Standardization R 1323565.1.012-2017]. Moscow, Standartinform Publ., 2017. (in Russian)
2. Ob utverzhdenii Trebovanij k sredstvam elektronnoj podpisi i Trebovanij k sredstvam udostoveryayushchego centra [On Approval of the Requirements for Cryptographic Modules for Digital Signature and Certificate Authority]. Order of the Federal Security Service of the Russian Federation of December 27, 2011 No. 796. (in Russian)
3. *Zima V. M., Kljuev A. V., Litvinov O. A., et al.* Osnovy zashchity informatsii ot nesanktsionirovannogo dostupa v avtomatizirovannykh sistemakh konfidentsial'nogo deloproizvodstva [Basics of protection of the information from unauthorized access in the automated systems of confidential office-work]. Tr. SPIIRAN, 2006, iss. 3, vol. 2, pp. 84–95. (in Russian)
4. *Khati L.* Full Disk Encryption and Beyond. Diss. Cryptography and Security [cs.CR], Universite PSL; ENS Paris — Ecole Normale Superieure de Paris, 2019. 182 p.
5. *Broz M.* Authenticated and Resilient Disk Encryption. PhD thesis, Brno, Masaryk University, 2018.
6. *Damgard I. and Dupont K.* Universally Composable Disk Encryption Schemes. IACR Cryptology ePrint Archive, 2005. https://eprint.iacr.org/2005/333.pdf.
7. *Gjosteen K.* Security notions for disk encryption. LNCS, 2005, vol. 3679, pp. 455–474.
8. https://integralmemory.com.

9. *Akhmetzyanova L., Alekseev E., Karpunin G., et al.* Ob odnom podkhode k formalizatsii zadach kriptograficheskogo analiza [On one approach to formalizing cryptographic analysis tasks]. Matem. Vopr. Kriptogr., 2020, to be published. (in Russian)
10. *Akhmetzyanova L., Alekseev E., Karpunin G., et al.* Chto plokhogo mozhno sdelat', nepravil'no ispol'zuya kriptoalgoritmy? [What can be done wrong by using cryptographic algorithms incorrectly?]. CTCrypt'2019. `https://ctcrypt.ru/files/files/2019/materials/29_Alekseyev.pdf`. (in Russian)
11. *Bhargavan K. and Leurent G.* On the practical (in-)security of 64-bit block ciphers. Collision attacks on HTTP over TLS and OpenVPN. Proc. CCS'16, October 24–28, 2016, Vienna, Austria, pp. 456–467. `https://sweet32.info/SWEET32_CCS16.pdf`.
12. *Smyshlyaev S.* Re-keying Mechanisms for Symmetric Keys. RFC 8645, August 2019. `https://tools.ietf.org/html/rfc8645`.
13. *Akhmetzyanova L. R., Alekseev E. K., Oshkin I. B., and Smyshlyaev S. V.* Increasing the Lifetime of Symmetric Keys for the GCM Mode by Internal Re-keying. Cryptology ePrint Archive: Report 2017/697.
14. *Alekseev E. K., Akhmetzyanova L. R., Meshkov D. A., et al.* O nagruzke na klyuch. Ch. 1 [On Key Lifetime. P. 1]. CRYPTO-PRO LLC Blog, 2017. `http://cryptopro.ru/blog/2017/05/17/o-nagruzke-na-klyuch-chast-1`. (in Russian)
15. *Alekseev E. K., Akhmetzyanova L. R., Meshkov D. A., et al.* O nagruzke na klyuch. Ch. 2 [On Key Lifetime. P. 2]. CRYPTO-PRO LLC Blog, 2017. `http://cryptopro.ru/blog/2017/05/29/o-nagruzke-na-klyuch-chast-2`. (in Russian)
16. *Bellare M. and Rogaway P.* Introduction to Modern Cryptography. 2005. `https://web.cs.ucdavis.edu/~rogaway/classes/227/spring05/book/main.pdf`.
17. *Bellare M., Boldyreva A., and O'Neill A.* Deterministic and efficiently searchable encryption. LNCS, 2007, vol. 4622, pp. 535–552.

# МАТЕМАТИЧЕСКИЕ ОСНОВЫ НАДЁЖНОСТИ ВЫЧИСЛИТЕЛЬНЫХ И УПРАВЛЯЮЩИХ СИСТЕМ

УДК 519.718

## О НАДЁЖНОСТИ СХЕМ ВО ВСЕХ ПОЛНЫХ БАЗИСАХ ИЗ ТРЁХВХОДОВЫХ ЭЛЕМЕНТОВ ПРИ НЕИСПРАВНОСТЯХ ТИПА 0 НА ВЫХОДАХ ЭЛЕМЕНТОВ

М. А. Алехина

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

Рассматривается реализация булевых функций схемами из ненадёжных функциональных элементов в полном базисе, содержащем функции трёх переменных. Предполагается, что элементы схемы переходят в неисправные состояния независимо друг от друга, подвержены однотипным константным неисправностям типа 0 на выходах. Для каждого полного базиса найдено либо точное значение коэффициента ненадёжности, либо его верхняя оценка.

**Ключевые слова:** *ненадёжные функциональные элементы, надёжность и ненадёжность схемы, синтез схем из ненадёжных элементов.*

DOI 10.17223/20710410/49/7

## ABOUT THE RELIABILITY OF LOGIC CIRCUITS IN ALL COMPLETE BASES WITH THREE-INPUT ELEMENTS AND FAILURES OF ZERO TYPE ON THEIR OUTPUTS

M. A. Alekhina

*Penza State Technological University, Penza, Russia*

**E-mail:** alekhina@penzgtu.ru

We consider the implementation of Boolean functions by circuits from unreliable functional elements in a complete basis containing functions of three variables. We suppose that the elements of the circuit pass to faulty states independently of each other, and they subject to the single-type constant faults of 0 type at outputs. For each complete basis, either the exact value of the coefficient of unreliability is found, or the upper estimate for this coefficient is calculated.

**Keywords:** *unreliable functional elements, reliability and unreliability of circuit, synthesis of circuits composed of unreliable elements.*

### Введение

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

К числу основных модельных объектов математической теории синтеза, сложности и надёжности управляющих систем относятся схемы из ненадёжных функциональных элементов, реализующие булевы функции. Проблема построения оптимальных по критериям надёжности и сложности схем из ненадёжных элементов является одной из наиболее важных и в то же время трудных в теории синтеза управляющих систем. Разработка специальных методов синтеза схем из ненадёжных функциональных элементов связана, главным образом, с выбранной математической моделью неисправностей. К основным моделям неисправностей относятся, например, инверсные и константные неисправности на выходах элементов. В работе рассматривается задача построения асимптотически оптимальных по надёжности схем в предположении, что функциональные элементы подвержены неисправностям типа 0 на выходах.

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

Любой метод синтеза схем из ненадёжных элементов характеризуется двумя важными параметрами: вероятностью ошибки на выходе схемы (ненадёжностью) и сложностью схемы. Именно минимизации сложности схем, реализующих булевы функции, уделено главное внимание в работах Р. Л. Добрушина, С. И. Ортюкова [2, 3], Д. Улига [4] и некоторых других авторов. Задача построения асимптотически оптимальных по надёжности схем из ненадёжных элементов, подверженных тем или иным неисправностям, ни Дж. фон Нейманом, ни другими исследователями до появления работ М. А. Алехиной не рассматривалась.

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

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

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

Позднее в работах В. В. Чугуновой, А. В. Васина, Д. М. Клянчиной и некоторых других авторов решалась задача реализации булевых функций асимптотически оптимальными по надёжности схемами при различных неисправностях элементов.

Эта работа продолжает исследования, начатые в [8] для полного базиса $\{\bar{x}_1 \vee \bar{x}_2\}$, элементы которого независимо друг от друга с вероятностью $\varepsilon$ подвержены неисправностям типа 0 на выходах. Позднее задача синтеза асимптотически оптимальных по надёжности схем при неисправностях типа 0 была решена [9] во всех полных неприводимых базисах из двухвходовых функциональных элементов, кроме одного. В этой работе исследуются полные базисы, содержащие функции трёх переменных. Для каждого из них найдено либо точное значение коэффициента ненадёжности, либо его верхняя оценка.

Аналогичная задача при инверсных неисправностях на выходах элементов решена А. В. Васиным [10] во всех полных базисах, содержащих функции трёх переменных, причём им найдены не только оценки коэффициента ненадёжности базиса, но и их точные значения.

Ранее [11 – 16] при неисправностях типа 0 на выходах элементов найдены функции трёх переменных (обозначим их множество через $G$) или пары функций, наличие которых в базисе гарантирует реализацию в этом базисе почти любой булевой функции схемой с ненадёжностью, асимптотически равной $\varepsilon$ при $\varepsilon \to 0$ (такие базисы имеют коэффициент ненадёжности 1).

Здесь исследованы все остальные полные базисы, содержащие функции трёх переменных, при неисправностях типа 0 на выходах элементов. В каждом из этих базисов найдены верхние оценки ненадёжности схем.

## 1. Необходимые понятия, определения и ранее известные результаты

Рассмотрим реализацию булевых функций схемами из ненадёжных функциональных элементов в полном конечном базисе $B$. Схема реализует функцию $f(x_1, \ldots, x_n)$, $n \in \mathbb{N}$, если при поступлении на входы схемы набора $\tilde{a}^n = (a_1, \ldots, a_n)$ при отсутствии неисправностей на выходе схемы появляется значение $f(\tilde{a}^n)$. Предполагается, что все функциональные элементы независимо друг от друга с вероятностью $\varepsilon$, $\varepsilon \in (0, 1/2)$, переходят в неисправные состояния типа 0 на выходах элементов. Эти неисправности характеризуются тем, что в исправном состоянии функциональный элемент реализует приписанную ему функцию, а в неисправном — константу 0.

Пусть $P_{\bar{f}(\tilde{a}^n)}(S, \tilde{a}^n)$ — вероятность появления значения $\bar{f}(\tilde{a}^n)$ на выходе схемы $S$, реализующей функцию $f(\tilde{x}^n)$ при входном наборе $\tilde{a}^n$. Ненадёжность схемы $S$ равна $P(S) = \max\{P_{\bar{f}(\tilde{a}^n)}(S, \tilde{a}^n)\}$, где максимум берётся по всем наборам $\tilde{a}^n$. Надёжность схемы $S$ равна $1 - P(S)$.

Пусть $P_\varepsilon(f) = \inf P(S)$, где инфимум берется по всем схемам $S$ из ненадёжных элементов, реализующим функцию $f(\tilde{x}^n)$ в базисе $B$. Схему $A$ из ненадёжных элементов, реализующую $f$, назовём *асимптотически оптимальной* (*асимптотически наилучшей*) *по надёжности*, если $P(A) \sim P_\varepsilon(f)$ при $\varepsilon \to 0$, т. е. $\lim\limits_{\varepsilon \to 0} \dfrac{P_\varepsilon(f)}{P(A)} = 1$.

**Замечание 1.** При неисправностях типа 0 на выходах элементов любая схема, содержащая хотя бы один функциональный элемент и реализующая отличную от константы 0 функцию, имеет ненадёжность не меньше $\varepsilon$ при всех $\varepsilon \in (0, 1/2)$ [9].

Полный конечный базис $B$ будем называть *базисом с коэффициентом ненадёжности* $k$ ($k \in \mathbf{N}$), если в этом базисе любую функцию можно реализовать схемой, функционирующей с ненадёжностью асимптотически не больше $k\varepsilon$ при $\varepsilon \to 0$ и найдется

функция, которую нельзя реализовать схемой, функционирующей с ненадёжностью асимптотически меньшей $k\varepsilon$.

Для неисправностей типа 0 на выходах элементов доказана следующая

**Теорема 1** [17]. В произвольном полном конечном базисе $B$ любую булеву функцию можно реализовать такой схемой $S$, что $P(S) \leqslant 3\varepsilon + 27\varepsilon^2$ при всех $\varepsilon \in (0, 1/960]$.

Из теоремы 1 следует, что при неисправностях типа 0 на выходах элементов любой полный конечный базис $B$ имеет коэффициент ненадёжности $k_B \in \{1, 2, 3\}$.

Пусть $P_2(n)$ — множество всех булевых функций, зависящих от $n$ переменных $x_1, x_2, \ldots, x_n$. Тогда $P_2(3)$ — множество всех булевых функций, зависящих от переменных $x_1, x_2, x_3$.

Булевы функции $f_1$ и $f_2$ назовём *конгруэнтными*, если одна из них может быть получена из другой заменой переменных (без отождествления). Пусть $X \subseteq P_2(3)$. Введём обозначение $\mathrm{Congr}\, X$ — множество всех функций, зависящих от переменных $x_1$, $x_2$, $x_3$, каждая из которых конгруэнтна некоторой функции множества $X$. Например, $\mathrm{Congr}\{1, x_1, x_1\&x_2\} = \{1, x_1, x_2, x_3, x_1\&x_2, x_2\&x_3, x_1\&x_3\}$. Обозначим:

$G_1 = \mathrm{Congr}\{x_1^{\sigma_1}x_2^{\sigma_2} \vee x_1^{\sigma_1}x_3^{\sigma_3} \vee x_2^{\sigma_2}x_3^{\sigma_3} : \sigma_i \in \{0,1\}, i \in \{1,2,3\}\}$, $|G_1| = 8$;

$G_2 = \mathrm{Congr}\{x_1^{\sigma_1}x_2^{\sigma_2} \oplus x_3^{\sigma_3} : \sigma_i \in \{0,1\}, i \in \{1,2,3\}\}$, $|G_2| = 24$;

$G_3 = \mathrm{Congr}\{x_1^{\sigma_1}x_2^{\bar\sigma_2} \vee x_2^{\sigma_2}x_3^{\sigma_3} : \sigma_i \in \{0,1\}, i \in \{1,2,3\}\}$, $|G_3| = 24$;

$G_4 = \mathrm{Congr}\{\bar x_1\bar x_2, \bar x_1\bar x_2\bar x_3\}$, $|G_4| = 4$;

$G_5 = \mathrm{Congr}\{x_1 \oplus x_2 \oplus a, x_1 \oplus x_2 \oplus x_3 \oplus b : a, b \in \{0,1\}\}$, $|G_5| = 8$;

$G_6 = \mathrm{Congr}\{\bar x_1(x_2 \oplus x_3 \oplus a) : a \in \{0,1\}\}$, $|G_6| = 6$;

$G_7 = \bigcup\limits_{i=1}^{6} G_i$, $|G_7| = 74$;

$G_8 = \mathrm{Congr}\{x_1^{\sigma_1}x_2^{\sigma_2} \vee x_1^{\bar\sigma_1}x_2^{\bar\sigma_2}x_3^{\bar\sigma_3} : \sigma_i \in \{0,1\}, i \in \{1,2,3\}\}$, $|G_8| = 24$;

$G_9 = \mathrm{Congr}\{x_1^{\sigma_1}x_2^{\sigma_2}x_3^{\sigma_3} \vee x_1^{\sigma_1}x_2^{\bar\sigma_2}x_3^{\bar\sigma_3} \vee x_1^{\bar\sigma_1}x_2^{\sigma_2}x_3^{\bar\sigma_3} : \sigma_i \in \{0,1\}, i \in \{1,2,3\}\}$, $|G_9| = 8$;

$G_{10} = \mathrm{Congr}\{\bar x_1(x_2^{\sigma_2} \vee x_3^{\sigma_3}) : \sigma_i \in \{0,1\}, i \in \{2,3\}\}$, $|G_{10}| = 12$;

$G_{11} = \mathrm{Congr}\{x_1^{\sigma_1}x_2^{\sigma_2}x_3^{\sigma_3} \vee x_1^{\bar\sigma_1}x_2^{\bar\sigma_2}x_3^{\bar\sigma_3} : \sigma_i \in \{0,1\}, i \in \{1,2,3\}\}$, $|G_{11}| = 4$;

$G_i^*$ — множество всех функций, двойственных функциям множества $G_i$ соответственно, $i \in \{8, 9, 10, 11\}$; $|G_i^*| = |G_i|$;

$G_{12} = \bigcup\limits_{i=8}^{11} (G_i \cup G_i^*)$, $|G_{12}| = 96$;

$G = G_7 \cup G_{12}$, $|G| = 170$.

Ранее для неисправностей типа 0 на выходах элементов в работах [11 – 15] получены результаты, которые можно сформулировать в виде теоремы 2.

**Теорема 2** [11 – 15]. Пусть полный конечный базис $B$ содержит функцию из множества $G$. Тогда любую булеву функцию в этом базисе можно реализовать такой схемой $S$, что $P(S) \leqslant \varepsilon + 100\varepsilon^2$ при всех $\varepsilon \in (0, 1/960]$.

В теореме 2, учитывая замечание 1, найдены базисы с коэффициентом ненадёжности 1.

Число функций в множестве $G$ равно 170 и составляет примерно 0,664 от числа всех функций из $P_2(3)$. Введём множества, содержащие остальные 86 функций из $P_2(3)$:

$\Theta = \mathrm{Congr}\{x_1x_2^{\sigma_2},\ x_1x_2^{\sigma_2}x_3^{\sigma_3},\ x_1(x_2 \oplus x_3)^{\sigma_1},\ x_1(x_2^{\sigma_2} \vee x_3^{\sigma_3}) : \sigma_i \in \{0,1\}, i \in \{1,2,3\}\}$, $|\Theta| = 34$;

$\Theta_1^* = \Theta^* \cup \mathrm{Congr}\{\bar x_1 \vee (x_2 \sim x_3)\}$, где $\Theta^*$ — множество функций, каждая из которых двойственна некоторой функции множества $\Theta$, $|\Theta_1^*| = 37$;

$\Omega = \mathrm{Congr}\{\bar x_1 \vee (x_2 \oplus x_3),\ \bar x_1 \vee \bar x_2,\ \bar x_1 \vee \bar x_2 \vee \bar x_3\}$, $|\Omega| = 7$;

$\Lambda = \mathrm{Congr}\{0, 1, x_1, \bar x_1\}$, $|\Lambda| = 8$.

Таким образом, множества $G, \Theta, \Theta_1^*, \Omega, \Lambda$ содержат все 256 булевых функций от переменных $x_1, x_2, x_3$, т. е. $G \cup \Theta \cup \Theta_1^* \cup \Omega \cup \Lambda = P_2(3)$.

**Теорема 3** [16]. Пусть полный базис $B$ таков, что $B \cap \Theta \neq \varnothing$ и $B \cap \Theta_1^* \neq \varnothing$. Тогда любую булеву функцию в этом базисе можно реализовать схемой, ненадёжность которой не больше $\varepsilon + 24\varepsilon^2$ при всех $\varepsilon \in (0, 1/960]$.

В теореме 3, учитывая замечание 1, также найдены базисы с коэффициентом ненадёжности 1, но в отличие от теоремы 2 здесь требуется, чтобы базис содержал сразу две функции определённого вида.

**Замечание 2.** Множество $\Theta$ сохраняет константу 0; множество $\Theta_1^*$ сохраняет константу 1.

Далее будем исследовать полные базисы $B \subseteq P_2(3) \setminus G$, для которых условия теоремы 3 не выполнены. Для доказательства новых результатов приведём следующую лемму:

**Лемма 1** [10]. Пусть $\psi \in \Theta_1^*$. Тогда подстановкой переменных из $\psi$ можно получить функцию вида $x_1 \vee x_2^b$, $b \in \{0, 1\}$.

## 2. Основные результаты

**Теорема 4.** Пусть полный базис $B$ таков, что $B \cap \Omega \neq \varnothing$. Тогда любую булеву функцию в этом базисе можно реализовать схемой, ненадёжность которой не больше $2\varepsilon + 24\varepsilon^2$ при всех $\varepsilon \in (0, 1/70]$.

***Доказательство.*** Покажем, что из каждой функции множества $\Omega$ подстановкой переменных можно получить функцию $\bar{x}_1 \vee \bar{x}_2$:

1) Пусть $\phi(x_1, x_2, x_3) = \bar{x}_1 \vee (x_2 \oplus x_3)$. Тогда $\phi(x_1, x_2, x_1) = \bar{x}_1 \vee (x_2 \oplus x_1) = \bar{x}_1 \vee \vee\, \bar{x}_1 x_2 \vee x_1 \bar{x}_2 = \bar{x}_1 \vee x_1 \bar{x}_2 = \bar{x}_1 \vee \bar{x}_2$.
2) Пусть $\phi(x_1, x_2, x_3) = \bar{x}_1 \vee \bar{x}_2 \vee \bar{x}_3$. Тогда $\phi(x_1, x_2, x_1) = \bar{x}_1 \vee \bar{x}_2$.

С учётом результатов [9] теорема доказана. ■

**Теорема 5.** Пусть $B$ — полный базис, $B \subseteq (\Theta_1^* \cup \Lambda)$. Тогда любую булеву функцию в этом базисе можно реализовать схемой, ненадёжность которой не больше $2\varepsilon + 42\varepsilon^2$ при всех $\varepsilon \in (0, 1/140]$.

***Доказательство.*** Пусть базис $B$ содержит некоторую функцию $\psi$ из множества $\Theta_1^*$. По лемме 1 из функции $\psi$ подстановкой переменных можно получить функцию вида $x_1 \vee x_2^b$, $b \in \{0, 1\}$.

Поскольку множество $\Theta_1^*$ сохраняет константу 1 (замечание 2), а по условию базис $B$ полный, то $B$ содержит функцию из $\Lambda$, которая не сохраняет константу 1. Таких функций две: $\bar{x}_1$ и константа 0. Возможны два варианта для параметра $b$:

1) $b = 0$, имеем функцию $x_1 \vee \bar{x}_2$, которая сохраняет константу 1. Кроме того, базис $B$ содержит хотя бы одну из функций: константу 0 или $\bar{x}_1$. В каждом из этих случаев теорема верна [9];
2) $b = 1$, имеем функцию $x_1 \vee x_2$. Кроме того, $B$ содержит $\bar{x}_1$ или 0. В первом случае (когда имеем $\bar{x}_1$), ссылаясь на [9], считаем теорему доказанной.
   Рассмотрим второй случай, т. е. из $\psi$ подстановкой переменных получили $x_1 \vee x_2$ и есть константа 0. Обе эти функции сохраняют константу 0, поэтому полный базис $B$ содержит не сохраняющую константу 0 функцию $\varphi \in \Theta_1^*$ (очевидно, что $\varphi$ и $\psi$ различны). Множество таких функций $\varphi$ с точностью до конгруэнтности — это $A = \{x_1 \vee \bar{x}_2,\ x_1 \vee \bar{x}_2 \vee x_3,\ x_1 \vee \bar{x}_2 \vee \bar{x}_3,\ x_1 \vee (x_2 \sim x_3),\ x_1 \vee \bar{x}_2 \bar{x}_3,\ \bar{x}_1 \vee (x_2 \sim x_3)\} \subset \Theta_1^*$. Нетрудно проверить, что из каждой функции множества $A$

подстановкой переменных можно получить функцию $x_1 \vee \bar{x}_2$, а этот случай рассмотрен в п. 1 доказательства.

Теорема доказана. ■

Введём два множества $\Theta_2$ и $\Theta_3$, которые являются подмножествами множества $\Theta$:

$\Theta_2 = \mathrm{Congr}\{x_1\bar{x}_2\bar{x}_3,\ x_1(x_2 \oplus x_3)^{\sigma_1} : \sigma_1 \in \{0,1\}\}$;

$\Theta_3 = \mathrm{Congr}\{x_1\bar{x}_2,\ x_1x_2\bar{x}_3,\ x_1(x_2 \vee \bar{x}_3),\ x_1(\bar{x}_2 \vee \bar{x}_3)\}$.

**Теорема 6.** Пусть $B$ — полный базис, $B \subseteq (\Theta_2 \cup \Lambda)$. Тогда любую булеву функцию в этом базисе можно реализовать схемой, ненадёжность которой не больше $2\varepsilon + 400\varepsilon^2$ при всех $\varepsilon \in (0, 1/1920]$.

***Доказательство.*** Поскольку базис $B$ полный, он содержит некоторую функцию $\psi$ из множества $\Theta_2$. Так как множество $\Theta_2$ сохраняет константу 0 (замечание 2), то базис $B$ содержит хотя бы одну функцию, которая не сохраняет константу 0. Таких функций в множестве $\Lambda$ две: $\bar{x}_1$ и константа 1. Следовательно, имеем четыре случая:

1) Базис $B$ содержит функции $x_1\bar{x}_2\bar{x}_3$ и $\bar{x}_1$. Отождествим переменные $x_2$ и $x_3$ и получим функцию $x_1\bar{x}_2$. В базисе $\{x_1\bar{x}_2, \bar{x}_1\}$ при $\varepsilon \in (0, 1/160]$ теорема верна [9].
2) Базис $B$ содержит функции $x_1\bar{x}_2\bar{x}_3$ и 1. Подставим константу 1 вместо переменной $x_1$ и получим функцию $\bar{x}_2\bar{x}_3$. Поскольку для ее реализации в рассматриваемом базисе требуется два элемента, в оценке $\varepsilon + 3\varepsilon^2$ ($\varepsilon \leqslant 1/160$), доказанной в [9], следует заменить $\varepsilon$ на $2\varepsilon$. В результате получим оценку $2\varepsilon + 12\varepsilon^2$ при $\varepsilon \leqslant 1/320$, теорема верна.
3) Базис $B$ содержит функции $x_1(x_2 \oplus x_3)^{\sigma_1}$ и $\bar{x}_1$. Подставим $\bar{x}_1$ вместо переменной $x_1$ и получим функцию $\bar{x}_1(x_2 \oplus x_3)^{\sigma_1} \in G_6$. Поскольку для её реализации в рассматриваемом базисе требуется два элемента, в оценке $\varepsilon + 100\varepsilon^2$ ($\varepsilon \leqslant 1/960$) из теоремы 2 следует заменить $\varepsilon$ на $2\varepsilon$. В результате получим оценку $2\varepsilon + 400\varepsilon^2$ при $\varepsilon \leqslant 1/1920$, теорема верна.
4) Базис $B$ содержит функции $x_1(x_2 \oplus x_3)^{\sigma_1}$ и 1. Подставим 1 вместо переменной $x_1$ и получим функцию $(x_2 \oplus x_3)^{\sigma_1} \in G_5$. Поскольку для её реализации в рассматриваемом базисе требуется два элемента, в оценке $\varepsilon + 100\varepsilon^2$ ($\varepsilon \leqslant 1/960$) из теоремы 2 следует заменить $\varepsilon$ на $2\varepsilon$. В результате получим оценку $2\varepsilon + 400\varepsilon^2$ при $\varepsilon \leqslant 1/1920$, теорема верна.

Теорема доказана. ■

**Теорема 7.** Пусть полный базис $B$ содержит функцию $\bar{x}_1$ и $B \cap \Theta_3 \neq \varnothing$. Тогда любую булеву функцию в этом базисе можно реализовать схемой, ненадёжность которой не больше $2\varepsilon + 12\varepsilon^2$ при всех $\varepsilon \in (0, 1/360]$.

***Доказательство.*** Поскольку $B \cap \Theta_3 \neq \varnothing$, имеем четыре случая:

1) Базис $B$ содержит функции $x_1\bar{x}_2$ и $\bar{x}_1$. В базисе $\{x_1\bar{x}_2, \bar{x}_1\}$ при $\varepsilon \in (0, 1/360]$ теорема верна [9].
2) Базис $B$ содержит функции $x_1x_2\bar{x}_3$ и $\bar{x}_1$. Отождествим переменные $x_1$ и $x_2$ и получим функцию $x_1\bar{x}_3$. В базисе $\{x_1\bar{x}_2, \bar{x}_1\}$ при $\varepsilon \in (0, 1/360]$ теорема верна [9].
3) Базис $B$ содержит функции $x_1(x_2 \vee \bar{x}_3)$ и $\bar{x}_1$. Подставим $\bar{x}_1$ вместо переменной $x_1$, затем отождествим переменные $x_1$ и $x_2$ и получим функцию $\bar{x}_1\bar{x}_3$. Поскольку для её реализации в рассматриваемом базисе требуется два элемента, в оценке $\varepsilon + 3\varepsilon^2$ ($\varepsilon \leqslant 1/160$), доказанной в [9], следует заменить $\varepsilon$ на $2\varepsilon$. В результате получим оценку $2\varepsilon + 12\varepsilon^2$ при $\varepsilon \leqslant 1/320$, теорема верна.
4) Базис $B$ содержит функции $x_1(\bar{x}_2 \vee \bar{x}_3)$ и $\bar{x}_1$. Подставим $\bar{x}_1$ вместо переменной $x_1$, затем отождествим переменные $x_2$ и $x_3$ и получим функцию $\bar{x}_1\bar{x}_3$. Поскольку

для её реализации в рассматриваемом базисе требуется два элемента, в оценке $\varepsilon + 3\varepsilon^2$ ($\varepsilon \leqslant 1/160$), доказанной в [9], следует заменить $\varepsilon$ на $2\varepsilon$. В результате получим оценку $2\varepsilon + 12\varepsilon^2$ при $\varepsilon \leqslant 1/320$, теорема верна.

Теорема доказана. ■

**Замечание 3.** Перечислим (с точностью до конгруэнтных функций) полные неприводимые базисы $B \subseteq ((\Theta \backslash \Theta_2) \cup \Lambda)$, не упомянутые в теореме 7: 1) $\{x_1x_2, \bar{x}_1\}$; 2) $\{x_1\bar{x}_2, 1\}$; 3) $\{x_1x_2x_3, \bar{x}_1\}$; 4) $\{x_1x_2\bar{x}_3, 1\}$; 5) $\{x_1(x_2 \vee x_3), \bar{x}_1\}$; 6) $\{x_1(\bar{x}_2 \vee \bar{x}_3), 1\}$.

**Замечание 4.** Для полных базисов $B \subseteq ((\Theta \backslash \Theta_2) \cup \Lambda)$, не упомянутых в теореме 7 (в том числе и неприводимых из замечания 3), выполняется теорема 1, а значит, коэффициент ненадёжности каждого из них не больше 3.

## 3. Выводы и рекомендации

В теоремах 4–7 исследованы полные базисы $B \subseteq P_2(3) \backslash G$, для которых условия теоремы 3 не выполняются. Доказано, что если базис $B$ удовлетворяет хотя бы одному из условий:

1) $B \cap \Omega \neq \varnothing$;
2) $B \subseteq (\Theta_1^* \cup \Lambda)$;
3) $B \subseteq (\Theta_2 \cup \Lambda)$;
4) полный базис $B$ содержит функцию $\bar{x}_1$ и $B \cap \Theta_3 \neq \varnothing$,

то любую булеву функцию можно реализовать схемой, функционирующей с ненадёжностью, асимптотически не больше $2\varepsilon$ при $\varepsilon \to 0$. Следовательно, коэффициент ненадёжности каждого из этих базисов не больше 2.

Учитывая результаты всех теорем 1–7, можно предложить следующий алгоритм получения оценки коэффициента ненадёжности $k_B$ полного базиса $B \subseteq P_2(3)$:

1. Найти $B \cap G$; если $B \cap G \neq \varnothing$, то $k_B = 1$, в противном случае перейти к шагу 2.
2. Найти $B \cap \Theta$ и $B \cap \Theta_1^*$. Если оба множества непустые, то $k_B = 1$, в противном случае перейти к шагу 3.
3. Проверить, верно ли хотя бы одно из условий:
   а) $B \cap \Omega \neq \varnothing$;
   б) $B \subseteq (\Theta_1^* \cup \Lambda)$;
   в) $B \subseteq (\Theta_2 \cup \Lambda)$;
   г) $B$ содержит функцию $\bar{x}_1$ и $B \cap \Theta_3 \neq \varnothing$.

   Если «да», то $k_B \leqslant 2$. Если «нет», то $k_B \leqslant 3$.

Таким образом, для произвольного полного базиса, содержащего функции трёх переменных, найдена верхняя оценка коэффициента ненадёжности.

Полученные в работе результаты справедливы в двойственных базисах при неисправностях типа 1 на выходах базисных элементов [18].

## ЛИТЕРАТУРА

1. *Фон Нейман Дж.* Вероятностная логика и синтез надежных организмов из ненадежных компонент // Автоматы. М.: ИЛ, 1956. С. 68–139.
2. *Добрушин Р. Л., Ортюков С. И.* Верхняя оценка для избыточности самокорректирующихся схем из ненадежных функциональных элементов // Проблемы передачи информации. 1977. Т. 13. № 3. С. 56–76.
3. *Ортюков С. И.* Об избыточности реализации булевых функций схемами из ненадежных элементов // Труды семинара по дискретной математике и её приложениям (Москва, 27–29 января 1987 г.). М.: Изд-во Моск. ун-та, 1989. С. 166–168.

4. *Uhlig D.* Reliable networks from unreliable gates with almost minimal complexity // LNCS. 1987. V. 278. P. 462–469.
5. *Pippenger N.* On networks of noisy gates // 26th Ann. Symp. Foundations of Computer Science. Portland, 21–23 Oct. 1985. P. 30–38.
6. *Яблонский С. В.* Асимптотически наилучший метод синтеза надежных схем из ненадежных элементов // Banach Center Publ. 1982. V. 7. No. 1. P. 11–19.
7. *Тарасов В. В.* К синтезу надежных схем из ненадежных элементов // Матем. заметки. 1976. Т. 20. № 3. С. 391–400.
8. *Алехина М. А.* О синтезе надежных схем из функциональных элементов $x/y$ при однотипных константных неисправностях на выходах элементов // Вест. Моск. ун-та. Матем. Механ. 1991. № 5. С. 80–83.
9. *Алехина М. А.* Синтез, надежность и сложность схем из ненадежных функциональных элементов: дис. ... докт. физ.-мат. наук. М.: МГУ им. М. В. Ломоносова, 2004.
10. *Васин А. В.* Асимптотически оптимальные по надежности схемы в полных базисах из трехвходовых элементов: дис. ... канд. физ.-мат. наук. Казань: Казанский (Приволжский) федеральный университет, 2010.
11. *Алехина М. А., Гусынина Ю. С., Шорникова Т. А.* О надежности схем при неисправностях типа 0 на выходах элементов в полном конечном базисе, содержащем особенную функцию // Изв. вузов. Математика. 2019. № 6. С. 85–88.
12. *Алехина М. А., Клянчина Д. М.* Об асимптотически оптимальных по надежности схемах в базисах, содержащих существенную линейную функцию и функцию вида $x_1^a \& x_2^b$ // Материалы XVI Междунар. конф. «Проблемы теоретической кибернетики» (Нижний Новгород, 20–25 июня 2011 г.). Н. Новгород: Изд-во Нижегород. ун-та, 2011. С. 33–37.
13. *Алехина М. А.* О надежности схем в полном конечном базисе, содержащем линейную функцию двух переменных и обобщенную дизъюнкцию // Известия высших учебных заведений. Поволжский регион. Физико-математические науки. 2019. № 1. С. 56–62.
14. *Alekhina M. A., Barsukova O. Yu., and Shornikova T. A.* On the reliability of circuits with type 0 faults at the outputs of the elements in the complete finite basis containing an essential linear function // Lobachevskii J. Mathematics. 2019. V. 40. No. 12. P. 2027–2033.
15. *Алехина М. А., Грабовская С. М., Гусынина Ю. С.* Достаточные условия реализации булевых функций асимптотически оптимальными по надежности схемами с тривиальной оценкой ненадежности при неисправностях типа 0 на выходах элементов // Прикладная дискретная математика. 2019. № 45. С. 44–54.
16. *Алехина М. А., Шорникова Т. А.* О надежности схем при неисправностях типа 0 на выходах элементов в полном конечном базисе, содержащем некоторые пары функций // Изв. вузов. Математика (в печати).
17. *Алехина М. А.* О надежности схем в произвольном полном конечном базисе при однотипных константных неисправностях на выходах элементов // Дискретная математика. 2012. Т. 24. Вып. 3. С. 17–24.
18. *Алехина М. А., Пичугина П. Г.* О надежности двойственных схем в полном конечном базисе // Материалы XVIII Междунар. школы-семинара «Синтез и сложность управляющих систем», Пенза, 28 сентября–3 октября 2009 г. М.: Изд-во мех.-мат. ф-та МГУ, 2009. С. 10–13.

## REFERENCES

1. *Von Neuman J.* Probabilistic logics and the synthesis of reliable organisms from unreliable components. Automata Studies / eds. C. Shannon and J. McCarthy. Princeton University Press, 1956, pp. 43–98.

2. *Dobrushin R. L. and Ortyukov S. I.* Upper bound on the redundancy of self-correcting arrangements of unreliable functional elements. Problems Inform. Transmission, 1977, vol. 13, iss. 3, pp. 203–218.
3. *Ortyukov S. I.* Ob izbytochnosti realizatsii bulevykh funktsiy skhemami iz nenadezhnykh elementov [On the redundancy of the Boolean functions implementation by circuits from unreliable elements]. Proc. Seminar Discr. Math. and its Appl. (Moscow, 27–29 Jan. 1987). Moscow, MSU Publ., 1989, pp. 166–168. (in Russian)
4. *Uhlig D.* Reliable networks from unreliable gates with almost minimal complexity. LNCS, 1987, vol. 278, pp. 462–469.
5. *Pippenger N.* On networks of noisy gates. 26th Ann. Symp. Foundations of Computer Science, Portland, 21–23 Oct. 1985, pp. 30–38.
6. *Yablonskiy C. V.* Asimptoticheski nailuchshiy metod sinteza nadezhnykh skhem iz nenadezhnykh elementov [Asymptotically best method for synthesizing reliable circuits from unreliable elements]. Banach Center Publ., 1982, vol. 7, no. 1, pp. 11–19. (in Russian)
7. *Tarasov V. V.* The synthesis of reliable circuits from unreliable elements. Math. Notes, 1976, vol. 20, iss. 3, pp. 775–780.
8. *Alekhina M. A.* O sinteze nadezhnykh skhem iz funktsional'nykh elementov $x/y$ pri odnotipnykh konstantnykh neispravnostyakh na vykhodakh elementov [On the synthesis of reliable circuits of $x/y$ functional elements at the same type constant faults at the element outputs]. Vestnic Moskovskogo Universiteta. Matematika. Mekhanika, 1991, no. 5, pp. 80–83. (in Russian)
9. *Alekhina M. A.* Sintez, nadegnost' i slognost' skhem iz nenadegnykh funktsional'nikh elementov [Synthesis, Reliability and Complexity of Circuits With Unreliable Functional Gates]. Doctoral dissertation in Mathematics and Physics, Penza, Penz. State Univ., 2004. 169 p. (in Russian)
10. *Vasin A. V.* Asimptoticheski optimal'nyye po nadezhnosti skhemy v polnykh bazisakh iz trekhvkhodovykh elementov [Asymptotically optimal on reliability circuits in complete bases of three-input elements]. PhD Thesis, Penza, Penz. State Univ., 2010. 100 p. (in Russian)
11. *Alekhina M. A., Gusynina Yu. S., and Shornikova T. A.* About reliability of circuits with faults of type 0 at the outputs of elements in a full finite basis containing a special function. Russian Mathematics, 2019, vol. 63, no. 6, pp. 79–81. (in Russian)
12. *Alekhina M. A. and Klyanchina D. M.* Ob asimptoticheski optimal'nykh po nadezhnosti skhemakh v bazisakh, soderzhashchikh sushchestvennuyu lineynuyu funktsiyu i funktsiyu vida $x_1^a \& x_2^b$ [On asymptotically optimal on reliability circuits in bases containing an essential linear function and a function of the form $x_1^a \& x_2^b$]. XVI Int. Conf. "Problems of theoretical cybernetics" (Nizhny Novgorod, 20–25 June 2011), Nizhny Novgorod, Nizhny Novgorod State Univ., 2011, pp. 33–37. (in Russian)
13. *Alekhina M. A.* O nadezhnosti skhem v polnom konechnom bazise, soderzhashchem lineynuyu funktsiyu dvukh peremennykh i obobshchennuyu diz"yunktsiyu [On the reliability of circuits in a complete finite basis containing a linear function of two variables and a generalized disjunction]. Izvestiya Vysshikh Uchebnykh Zavedeniy. Povolzhskiy Region. Fiziko-Matematicheskie Nauki, 2019, no. 1 (49), pp. 56–62. (in Russian)
14. *Alekhina M. A., Barsukova O. Yu., and Shornikova T. A.* On the reliability of circuits with type 0 faults at the outputs of the elements in the complete finite basis containing an essential linear function. Lobachevskii J. Mathematics, 2019, vol. 40, no. 12, pp. 2027–2033.
15. *Alekhina M. A., Grabovskaya S. M., and Gusynina Yu. S.* Dostatochnye usloviya realizatsii bulevykh funktsiy asimptoticheski optimalnymi po nadegnosti skhemami s trivial'noy otsenkoy nenadegnosti pri neispravnostyakh tipa 0 na vykhodakh elementov [Sufficient conditions for implementation of Boolean functions by asymptotically optimal on reliability

circuits with the trivial estimate of unreliability in the case of faults of type 0 at the element outputs]. Prikladnaya Diskretnaya Matematika, 2019, no. 45, pp. 44–54. (in Russian)

16. *Alekhina M. A. and Shornikova T. A.* O nadezhnosti skhem pri neispravnostyakh tipa 0 na vykhodakh elementov v polnom konechnom bazise, sodergashchem nekotoreye pary funktsiy [On the reliability of circuits with type 0 faults at the outputs of elements in a complete finite basis containing some pairs of functions]. Russian Mathematics. To be published. (in Russian)
17. *Alekhina M. A.* On reliability of circuits over an arbitrary complete finite basis under single-type constant faults at outputs of elements. Discr. Math. Appl., 2012, no. 22(4), pp. 383–391.
18. *Alekhina M. A. and Pichugina P. G.* O nadezhnosti dvoystvennykh skhem v polnom konechnom bazise [On the reliability of dual circuts in the complete finite basis]. XVIII Int. School-Seminar "Synthesis and complexity of control systems" (Penza, 28 Sept.–3 Oct. 2009), Moscow, Faculty of Mechanics and Mathematics of Moscow State University, 2009, pp. 10–13. (in Russian)

# ПРИКЛАДНАЯ ТЕОРИЯ ГРАФОВ

УДК 519.87

## ПАРАМЕТРИЧЕСКОЕ ЗАДАНИЕ СЕРИИ СЕМЕЙСТВ АНАЛИТИЧЕСКИ ОПИСЫВАЕМЫХ ЦИРКУЛЯНТНЫХ СЕТЕЙ СТЕПЕНИ ШЕСТЬ[1]

Э. А. Монахова

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

Получена серия семейств неориентированных кольцевых циркулянтных сетей степени шесть любого заданного диаметра $d > 1$, которая включает в том числе циркулянтные сети максимального порядка для всех диаметров $d \equiv 0 \pmod 3$ и $d \equiv 2 \pmod 3$. Серия семейств задаётся определяющими соотношениями между порядком графа и его образующими и порождающим параметром $p$, $1 \leqslant p < d$, при этом образующие и порядки графов являются полиномами третьей степени относительно диаметра графа. Приведены примеры построения новых семейств циркулянтных сетей степени шесть на основе задания функций $p = p(d)$.

**Ключевые слова:** *неориентированные циркулянтные сети степени шесть, циркулянтные графы заданного диаметра, семейства циркулянтных графов.*

DOI 10.17223/20710410/49/8

## A SET OF FAMILIES OF ANALYTICALLY DESCRIBED TRIPLE LOOP NETWORKS DEFINED BY A PARAMETER

E. A. Monakhova

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

**E-mail:** emilia@rav.sscc.ru

A set of families of undirected triple loop networks of the form $C(N(d,p); 1, s_2(d,p), s_3(d,p))$ with the given diameter $d > 1$ and a parameter $p = 1, 2, \ldots, d-1$ is obtained. For each such family, the order $N$ of every graph in the family and its generators $s_2$ and $s_3$ are defined by a cubical polynomial function of the diameter. The found set includes circulant graphs of degree 6 with the largest known orders for any diameters $d \equiv 0 \pmod 3$ and $d \equiv 2 \pmod 3$. Examples of constructing new families of triple loop networks based on the definition of functions $p = p(d)$ are presented.

**Keywords:** *undirected triple loop networks, circulant graphs of degree 6 with given diameter, families of circulant graphs.*

[1]Исследование выполнено в рамках бюджетного проекта ИВМиМГ СО РАН № 0315-2019-0006.

## Введение

Циркулянтные сети (графы) (см. обзоры [1 – 5]) широко изучаются в качестве популярной топологии для мультипроцессорных систем и компьютерных сетей и в ряде других приложений. Актуальным становится их применение в качестве топологии для сетей на кристалле (networks-on-chip) [6 – 8]. Это обусловлено их лучшими структурными характеристиками и высокими показателями масштабируемости при большом количестве узлов по сравнению со стандартными топологиями сетей на кристалле. Важной задачей для сетей на кристалле с циркулянтной топологией является разработка эффективных алгоритмов маршрутизации, связанных с особенностями требований, предъявляемых к используемым ресурсам сетей на кристалле.

Дадим определение циркулянтных сетей. Пусть $n, N \in \mathbb{N}$, $S = \{s_1, s_2, \ldots, s_n\}$ — множество целых чисел, $1 \leqslant s_1 < \ldots < s_n \leqslant \lfloor N/2 \rfloor$. Неориентированный *циркулянтный* граф $C(N; s_1, \ldots, s_n)$ имеет множество вершин $V = \mathbb{Z}_N = \{0, 1, \ldots, N-1\}$ и множество рёбер $A = \{(v, v \pm s_i \bmod N) : v \in V, i = 1, \ldots, n\}$. Числа $s_1, s_2, \ldots, s_n$ называются образующими графа, $N$ — его порядком, $n$ — размерностью, степень вершин графа равна $2n$. Будем исследовать кольцевые циркулянтные сети вида $C(N; 1, s_2, \ldots, s_n)$ с $s_1 = 1$, изучаемые в литературе как самостоятельный класс графов. Диаметр графа (оценивает максимальную структурную задержку в сети) равен $\mathrm{d}(C(N; S)) = \max\limits_{u,v \in V} \mathrm{d}(u, v)$, где $\mathrm{d}(u, v)$ — длина кратчайшего пути между вершинами $u$ и $v$.

В литературе известны следующие аналитически описываемые семейства кольцевых циркулянтных сетей степени шесть и диаметра $d$: графы вида $C(3d^2 + 3d + 1; 1, 3d + 1, 3d + 2)$ [9]; циркулянтные сети с $N = 8d^3/27 + 4d^2/3 + 2d + 1$ [10]; циркулянты с $N = 32\lfloor d/3 \rfloor^3 + 8\lfloor d/3 \rfloor^2 + 2\lfloor d/3 \rfloor$ [11] и диаметром, меньшим или равным $d$, где $d \geqslant 3$. Получены алгоритмы поиска кратчайших путей [12, 13] для семейства из [9] и эквивалентных графов вида $C(3d^2 + 3d + 1; d, d + 1, 2d + 1)$ [14]. В [15] найдено семейство трёхмерных циркулянтов с порядком $N = 4d^2 - 2d - 2$, где $d \equiv 3, 5 \pmod 6$ — диаметр графов, как решение оптимизационной задачи на максимум при рассмотрении произведения Кронекера двух циркулянтов степеней два и три. В [15] дан алгоритм поиска кратчайших путей для графов найденного семейства. Следует отметить, что образующие графа $C(3d^2 + 3d + 1; 1, 3d + 1, 3d + 2)$, где $d \geqslant 1$, получены в [9] как решение оптимизационной задачи на максимум при укладке (tessellation) трёхмерного графа на плоскости, когда рассматриваются графы диаметра $d$ вида $C(N_{\max} = 3d^2 + 3d + 1; s_1, s_2, s_1 + s_2)$. Заметим, что в большинстве этих работ порядки графов рассматриваемых семейств — это квадратичные функции от диаметра, хотя больший интерес представляет получение семейств с кубической функцией от диаметра, как более плотных и компактных графов. В [16] найдено семейство циркулянтных сетей степени шесть с максимальным порядком среди всех кольцевых циркулянтов заданного диаметра $d$ и приведён аналитический алгоритм поиска кратчайших путей для найденного семейства. В работе [8] для кольцевых циркулянтных сетей степени шесть общего вида предложены различные алгоритмы поиска кратчайших путей и даны оценки требуемых ресурсов при реализации в сетях на кристалле.

В настоящей работе представлено параметрически задаваемое аналитическое описание кольцевых циркулянтных графов степени шесть, которое порождает серию семейств циркулянтных сетей, включающую в том числе семейство графов с максимально возможным порядком для любого заданного диаметра, а также позволяет синтезировать новые семейства с лучшими структурными характеристиками, чем известные в литературе. Интересным приложением полученного результата является возмож-

ность решения проблемы поиска кратчайших путей в семействах циркулянтных сетей степени шесть с помощью аналитического метода, общего для всех графов семейства.

## 1. Теорема о построении серии циркулянтных графов степени шесть

Рассмотрим множество трёхмерных циркулянтных графов вида $C(N;1,s_2,s_3)$, где $1<s_2<s_3\leqslant\lfloor N/2\rfloor$. Будем использовать обозначение $D(x)$, $0\leqslant x<N$, для длины кратчайшего пути из вершины 0 в вершину $x$. В графе $C(N;1,s_2,s_3)$ выделим две ближайшие по циклу, образованному образующей $s_1$, вершины $u,v$, $u<v$, такие, что значения $D(u)$ и $D(v)$ получены без использования образующих $\pm s_1$. Тогда расстояния из вершины 0 до всех вершин, лежащих между $u$ и $v$, могут быть вычислены с использованием того факта, что разница между смежными вершинами равна единице.

**Лемма 1.** Пусть в циркулянтном графе $C(N;1,s_2,s_3)$ вершины $u,v$, $u<v$, — ближайшие по циклу, заданному образующей $s_1=1$, значения которых $D(u)$ и $D(v)$ получены без использования образующих $\pm1$. Тогда

$$\max_{u\leqslant x\leqslant v}D(x)=\lfloor(D(u)+D(v)+v-u)/2\rfloor \tag{1}$$

и достигается в вершине $x=\lfloor(v+u+D(v)-D(u))/2\rfloor$.

Следующая теорема даёт возможность построения целой серии семейств рассматриваемых графов заданного диаметра, что достигается введением в аналитическое описание графов параметра $p$, зависящего от диаметра. Эта теорема задаёт один из возможных типов определяющих соотношений между порядком графа и его образующими, когда и порядок графа $N$, и образующие $s_2$ и $s_3$ являются полиномами третьей степени относительно диаметра.

**Теорема 1.** Для каждого целого $d>1$ пусть

$$p=1,2,\ldots,d-1. \tag{2}$$

Тогда диаметр циркулянтных графов вида $C(N;1,s_2,s_3)$, где

$$\begin{cases} N=8p^3-(16d+8)p^2+(8d^2+8d)p+2d+1,\\ s_2=4p(d-p)^2+2p(d-p)+d-3p,\\ s_3=s_2+4p, \end{cases} \tag{3}$$

равен $d$.

***Доказательство.*** Рассмотрим циркулянтный граф $C(N;1,s_2,s_3)$ вида (3). Пусть

$$\begin{gathered}\Delta=s_3-s_2=4p,\\ r=(d-p)\Delta+\Delta/2+1.\end{gathered} \tag{4}$$

Согласно (3), порядок графа равен произведению двух нечётных чисел $N=(2(d-p)+1)r$ и его образующие имеют вид $s_2=(d-p)r-\Delta/2$, $s_3=(d-p)r+\Delta/2$.

Поскольку число вершин графа состоит из целого числа интервалов длины $r$, будем называть их $r_i$-интервалами на графе, где $0\leqslant i\leqslant 2(d-p)$ — номер интервала, или кратко $r_i=[ir,ir+r]$. Симметрия функции расстояний $D(x)$ в циркулянтах относительно $N/2$ позволяет в дальнейшем ограничиться значениями $i=0,1,\ldots,d-p$. Будем также использовать термин $\Delta$-интервалы для обозначения перемещений (прыжков) по $r_i$-интервалу на длину $\Delta$ из вершин левого или правого концов $r_i$-интервала. В силу (4)

перемещение на величину $\Delta$ даёт приращение функции расстояния $D(x)$ на 2. Так как $r = N-(s_2+s_3)$, перемещения в графе на величину $r$ также дают приращение функции расстояния $D(x)$ на 2.

Выделим в рассматриваемом графе две вершины $F$ и $R = N - F$, $F < N/2 < R$, играющие ключевую роль в определении функции расстояний:

$$F = (s_2 + s_3)/2 = (d - p)r.$$

Имеем $R = F + r$, $D(F \pm \Delta/2) = D(R \pm \Delta/2) = 1$. Тогда, выбирая минимальный из двух возможных путей в вершину $F$ (или $R$) из 0, получим

$$D(F) = D(R) = \begin{cases} 2p + 1 & \text{при} \quad 1 \leqslant p < \lceil d/2 \rceil, \\ 2(d - p) & \text{при} \quad \lceil d/2 \rceil \leqslant p \leqslant d - 1. \end{cases} \tag{5}$$

Для определения диаметра рассматриваемого графа учитываем следующее. Каждый из интервалов вида

$$[js_2 \bmod N, js_3 \bmod N],\ 1 \leqslant j \leqslant d,$$

состоит из $j$ $\Delta$-интервалов, на концах которых вершины $x$ имеют расстояния до нуля $D(x) = j$. В силу леммы 1 в серединах этих $\Delta$-интервалов находятся вершины с максимумами (равными) расстояний до 0. Аналогичные рассуждения применимы для интервалов вида

$$[N - js_3 \bmod N, N - js_2 \bmod N],\ 1 \leqslant j \leqslant d.$$

Согласно (3) и учитывая, что $N = 2F + r$, для $N$ выполняется условие

$$jF \equiv -2iF \pmod N$$

для нечётных $j$, где

$$2i + j = 2(d - p) + 1. \tag{6}$$

Рассмотрим $r_i$-интервал, где $i \in \{0, \ldots, d - p\}$. Имеем

$$-2is_2 \bmod N = i(r + \Delta), \quad -(2i + 2)s_3 \bmod N = (i + 1)(r - \Delta),$$
$$js_3 \bmod N = ir + j\Delta/2, \quad (j - 2)s_2 \bmod N = (i + 1)r - (j - 2)\Delta/2.$$

Таким образом,

$$D(i(r + \Delta)) = 2i, \quad D((i + 1)(r - \Delta)) = 2i + 2; \tag{7}$$
$$D(ir + j\Delta/2) = j, \quad D((i + 1)r - (j - 2)\Delta/2) = j - 2. \tag{8}$$

Итак, надо доказать, что для графов вида (3) максимальное расстояние до вершины 0 из любой вершины $x$, $0 \leqslant x \leqslant \lfloor N/2 \rfloor$, равно $d$. Для удобства представления будем рассматривать интервал $0 \leqslant x \leqslant R$. Всё множество вершин $\{0, \ldots, R\}$ разобъём на шесть подмножеств $V_m$, $0 \leqslant m \leqslant 5$, $\sum\limits_{m=0}^{5} |V_m| = (d - p + 1)r$, соответственно типам содержащихся в них $r_i$-интервалов (табл. 1 и 2). Тип $r_i$-интервала определяется значениями функции $D(x)$ на его концах (например, для множеств $V_0$, $V_1$ и $V_2$ имеем $D(ir) = 2i$,

Таблица 1

**Распределение вершин $0 \leqslant x \leqslant R$ графов вида (3) по типам $r_i$-интервалов при наличии обратной волны из $F$**

| Параметр $p$: $1 \leqslant p < \lceil d/2 \rceil$ | | | |
|---|---|---|---|
| Множество | Тип $r_i$-интервала | Значения $i$ | Мощность множеств |
| $V_0$ | $2i, 2i+2$ | $0 \leqslant i < \lfloor d/2 \rfloor - p$ | $\lvert V_0\rvert = (\lfloor d/2 \rfloor - p)r$ |
| $V_1$ | $2i, j-2, 2i+2$ | $i = \lfloor d/2 \rfloor - p$ | $\lvert V_1\rvert = r$ |
| $V_2$ | $2i, j, j-2, 2i+2$ | $\lfloor d/2 \rfloor - p < i < \lfloor d/2 \rfloor$ | $\lvert V_2\rvert = (p-1)r$ |
| $V_3$ | $2i, j, j-2, 2d-1-2i$ | $i = \lfloor d/2 \rfloor$ | $\lvert V_3\rvert = r$ |
| $V_4$ | $2d+1-2i, j, j-2, 2d-1-2i$ | $\lfloor d/2 \rfloor < i < d-p$ | $\lvert V_4\rvert = (\lceil d/2 \rceil - p - 1)r$ |
| $V_5$ | $2p+1, j, j, 2p+1$ | $i = d-p$ | $\lvert V_5\rvert = r$ |

Таблица 2

**Распределение вершин $0 \leqslant x \leqslant R$ графов вида (3) по типам $r_i$-интервалов при отсутствии обратной волны из $F$**

| Параметр $p$: $\lceil d/2 \rceil \leqslant p \leqslant d-1$ | | | |
|---|---|---|---|
| Множество | Тип $r_i$-интервала | Значения $i$ | Мощность множеств |
| $V_1$ | $2i, j-2, 2i+2$ | $i = 0,\ p = d/2$ | $\lvert V_1\rvert = \begin{cases} r, & p = d/2, \\ 0, & p \neq d/2 \end{cases}$ |
| $V_2$ | $2i, j, j-2, 2i+2$ | $\begin{cases} 0 < i < d-p, & p = d/2, \\ 0 \leqslant i < d-p, & p \neq d/2 \end{cases}$ | $\lvert V_2\rvert = \begin{cases} (p-1)r, & p = d/2, \\ (d-p)r, & p \neq d/2 \end{cases}$ |
| $V_5$ | $2(d-p), j, j, 2(d-p)$ | $i = d-p$ | $\lvert V_5\rvert = r$ |

$D(ir+r) = 2i+2$), а также тем, учитываются или нет значения $j$ и $j-2$ в определении функции расстояния $D(x)$ вершин внутри интервала (учитываются значения, не превышающие $d$). Для наглядности доказательства теоремы все основные параметры, относящиеся к множествам $V_m$, $0 \leqslant m \leqslant 5$, суммированы в табл. 1 и 2.

В табл. 1 представлены результаты, когда есть прямая волна расстояний, порождённая $r$-интервалами из 0, и есть обратная волна расстояний, порождённая $r$-интервалами из вершины $F$ (см. первое соотношение (5)). В табл. 1 выделяются особые случаи: $V_0 = \varnothing$ при $p = \lfloor d/2 \rfloor$, $V_2 = \varnothing$ при $p = 1$, $V_4 = \varnothing$ при $p = \lceil d/2 \rceil - 1$.

В табл. 2 представлены результаты, когда есть прямая волна расстояний, порождённая $r$-интервалами из 0, и нет обратной волны расстояний, порождённой $r$-интервалами из вершины $F$ (см. второе соотношение (5)). В этом случае в графе отсутствуют множества вершин типа $V_0$, $V_3$ и $V_4$, а также $V_1 = \varnothing$ при $p \geqslant (d+1)/2$.

В процессе доказательства далее будем разделять вершины, принадлежащие всем $r_i$-интервалам из множеств $V_m$, $0 \leqslant m \leqslant 5$, на три множества (интервала вершин): $r_i = A_1 \cup A_2 \cup A_3$, где $\sum\limits_{k=1}^{3} |A_k| = r$. Обозначив через $x_l$ ($x_r$) номера вершин в $A_2$, соответствующие левому (правому) концам интервала $A_2$, получим

$$A_1 = [ir, x_l],\ A_2 = [x_l, x_r],\ A_3 = [x_r, ir+r].$$

1) Пусть $x \in V_0$, где $1 \leqslant p < \lceil d/2 \rceil$, $p \neq \lfloor d/2 \rfloor$ (табл. 1).

Множество вершин $V_0$ состоит из $r_i$-интервалов, $0 \leqslant i < q = \lfloor d/2 \rfloor - p$, где $D(ir) = 2i$, $D(ir+r) = 2i+2$ и значения $j$, $j-2$ не учитываются при расчёте $D(x)$ для вершин $x \in r_i$. Для $V_0$ определим $x_l = ir + q\Delta$, $x_r = (i+1)r - q\Delta$, $|A_1| = |A_3| = q\Delta$.

Для $A_1$ имеем $D(x) = 2i$ на концах $i$ $\Delta$-интервалов, затем на концах оставшихся $(q-i)$ $\Delta$-интервалов значения $D(x)$ увеличиваются на 2, достигая значения $D(x_l) = 2q$.

Таким образом, для $A_1$ функция $D(x)$ достигает максимума, когда $i = q-1$. В силу (1) получаем $\max\limits_{x\in A_1} D(x) = 2\lfloor d/2\rfloor - 1 < d$.

Для $A_2$ имеем $D(x_l) = D(x_r) = 2(\lfloor d/2\rfloor - p)$. Из вершин $x_l$ и $x_r$ навстречу друг другу идут волны $\Delta$-интервалов, увеличивающих на 2 значения $D(x)$, которые заканчиваются, когда $D(x)$ достигает значений $2\lfloor d/2\rfloor$. В силу (1) получим для всех $i$ $\max\limits_{x\in A_2} D(x) = d$.

Учитывая для $r_i$-интервала, что, начиная с вершины $x = ir + r$, $D(x) = 2i + 2$ на концах $(i + 1)$ $\Delta$-интервалов, получаем, по аналогии с $A_1$, $\max\limits_{x\in A_3} D(x) = 2\lfloor d/2\rfloor \leqslant d$.

2) Пусть $x \in V_1$, где $1 \leqslant p < \lceil d/2\rceil$ и $p = d/2$.

Множество вершин $V_1$ состоит из одного $r_i$-интервала, где $i = \lfloor d/2\rfloor - p$ (см. табл. 1 и 2). Для него $D(ir) = 2i$, $D(ir + r) = 2i + 2$, и при расчёте $D(x)$ для вершин $x \in r_i$ учитываем только $j - 2 \leqslant d$. Определим $x_l = ir + i\Delta$, $x_r = (i + 1)(r - \Delta)$, $|A_1| = i\Delta$, $|A_3| = (i + 1)\Delta$.

Для $A_1$ имеем $D(x) = 2i$ на концах всех $i$ $\Delta$-интервалов. Таким образом, применяя (1), получим $\max\limits_{x\in A_1} D(x) = \lfloor(2i+2i+\Delta)/2\rfloor = 2\lfloor d/2\rfloor \leqslant d$. Отметим, что $A_1 = \varnothing$ при $p = d/2$.

Для $A_2$, согласно (7), имеем $D(x_l) = 2i$ и $D(x_r) = 2i+2$. Из вершин $x_l$ и $x_r$ навстречу идут волны $\Delta$-интервалов, увеличивающих на 2 значения $D(x)$. Волны заканчиваются, когда $D(x)$ достигает при нечётных $d$ значения $(d - 1)$ или при чётных $d$ — значения $(d-2)$ при движении из $x_l$ или значения $d$ при движении из $x_r$. Применяя (1), получим: $\max\limits_{x\in A_2} D(x) = \lfloor(2i + d + \Delta/2 + 1)/2\rfloor = d$ при чётных $d$; $\max\limits_{x\in A_2} D(x) = \max\{\lfloor(2i + 2i + 2 + + \Delta)/2\rfloor, \lfloor(2i + 2 + d - 1 + \Delta/2 + 1)/2\rfloor\} = d$ при нечётных $d$. Отметим, что и в случае $p = 1$ при чётных $d$, когда $|A_2| = 3$, $\max\limits_{x\in A_2} D(x) = d$.

Для $A_3$ имеем $D(x) = 2i + 2$ на концах всех $(i + 1)$ $\Delta$-интервалов, $D(x) = j - 2 = 2\lceil d/2\rceil - 1$ в их серединах. Таким образом, применяя (1) и (6), получаем $\max\limits_{x\in A_3} D(x) = \lfloor(2i + 2 + j - 2 + \Delta/2)/2\rfloor = \lfloor(2(d - p) + 1 + 2p)/2\rfloor = d$.

3) Пусть $x \in V_2$, где $1 \leqslant p \leqslant d - 1$.

Множество $V_2$ состоит из $r_i$-интервалов, где $D(ir) = 2i$, $D(ir + r) = 2i + 2$, и при расчёте $D(x)$ для вершин $x \in r_i$ учитываем значения $j$ и $j-2$. Значения $i$ представлены в табл. 1 и 2. Определение $x_l$ и $x_r$ на $r_i$-интервале зависит от значений $i$:

*а*) при $2i < d - p - 1$: $x_l = i(r + \Delta) + \Delta/2$, $x_r = (i + 1)(r - \Delta) - \Delta/2$, $|A_2| = = (d - p - 2 - 2i)\Delta + \Delta/2 + 1$;

*б*) при $2i = 2\lfloor d/2\rfloor - p$: $x_l = i(r + \Delta) + \Delta/2$, $x_r = (i + 1)r - \lfloor j/2\rfloor\Delta$, $|A_2| = 1$;

*в*) при $2i > d-p$: $x_l = ir+\lceil j/2\rceil\Delta$, $x_r = (i+1)r-\lfloor j/2\rfloor\Delta$, $|A_2| = (d-p-j)\Delta+\Delta/2+1$.

Для $A_1$ имеем $D(x) = 2i$ на концах $i$ $\Delta$-интервалов и $D(x) = j$ в их серединах во всех случаях *а*–*в*. Для $A_3$ аналогично имеем $D(x) = 2i + 2$ на концах $(i + 1)$ $\Delta$-интервалов, $D(x) = j - 2$ в их серединах. Таким образом, применяя (1) и (6), получаем $\max\limits_{x\in A_1} D(x) = \max\limits_{x\in A_3} D(x) = \lfloor(2i + j + \Delta/2)/2\rfloor = d$.

Для $A_2$, согласно (7), имеем $D(x_l - \Delta/2) = 2i$ и $D(x_r + \Delta/2) = 2i + 2$ в случае *а*; согласно (8), $D(x_l - \Delta/2) = j$ и $D(x_r + \Delta/2) = j - 2$ в случае *в*. Из вершин $(x_l - \Delta/2)$ и $(x_r + \Delta/2)$ на множестве $A_2$ навстречу друг другу идут волны $\Delta$-интервалов, увеличивающих на 2 значения $D(x)$. Учитывая значения $|A_2|$ и применяя (1), получим: $\max\limits_{x\in A_2} D(x) = \lfloor(2(2i + 2) + 2(d - p - 2i - 2) + \Delta/2 + 1)/2\rfloor = d$ в случае *а*; $\max\limits_{x\in A_2} D(x) = \lfloor(2j + 2(d - p - j) + \Delta/2 + 1)/2\rfloor = d$ — в случае *в*.

4) Пусть $x \in V_3$, где $1 \leqslant p < \lceil d/2 \rceil$.

Множество $V_3$ состоит из $r_i$-интервала, для которого $i = \lfloor d/2 \rfloor$, $D(ir) = 2i$, значение $D(ir+r) = 2\lceil d/2 \rceil - 1$ формируется обратной волной длины $r$ из вершины $F$. Поскольку $j = 2(\lceil d/2 \rceil - p) + 1 \leqslant d$, при расчёте $D(x)$ для вершин $x \in r_i$ учитываем $j$ и $j-2$. Определим $x_l = ir + j\Delta/2$, $x_r = (i+1)r - j\Delta/2$, $|A_1| = |A_3| = j\Delta/2$.

Для $A_1$ имеем $D(x) = 2i$ на концах $i$ $\Delta$-интервалов и $D(x) = j$ в их серединах. Таким образом, применяя (1) и (6) и учитывая значение $|A_1|$, получим $\max\limits_{x \in A_1} D(x) = \lfloor (2i + j + \Delta/2)/2 \rfloor = d$.

Так как в случае $p = 1$ при нечётных $d$ значение $|A_2| = -1$, то для него $A_2$ не рассматривается. При $p > 1$ для $A_2$ имеем $D(x_l) = D(x_r) = j$. Из вершин $x_l$ и $x_r$ навстречу идут волны $\Delta$-интервалов, увеличивающих на 2 значения $D(x)$, которые заканчиваются, когда $D(x)$ достигает значений $2\lceil d/2 \rceil - 1$. Из (1) следует $\max\limits_{x \in A_2} D(x) = \lfloor (2j + 2(d - p - j) + \Delta/2 + 1)/2 \rfloor = d$.

Для $A_3$ имеем $D(x_r) = j$, $D(ir+r) = 2\lceil d/2 \rceil - 1$ и $D(x) = j-2$ для остальных концов $\Delta$-интервалов. Применяя (1), получим $\max\limits_{x \in A_3} D(x) = \max\{2\lceil d/2 \rceil - 1, \lfloor (2(j-2) + \Delta)/2 \rfloor\} = d$.

5) Пусть $x \in V_4$, где $1 \leqslant p < \lceil d/2 \rceil$.

Множество $V_4$ состоит из $r_i$-интервалов, $\lfloor d/2 \rfloor < i < d - p$, для которых $D(ir) = (2d + 1 - 2i)$, $D(ir + r) = (2d - 1 - 2i)$ — нечётные значения, образованные обратной волной длины $r$ из вершины $F$. При расчёте $D(x)$ для вершин $x \in r_i$ учитываем $j$ и $j - 2$. Положив $x_l = ir + j\Delta/2$, $x_r = (i+1)r - j\Delta/2$, получим $|A_1| = |A_3| = j\Delta/2$.

Для $A_1$ имеем $D(x) = j$ на концах всех $\Delta$-интервалов, $D(ir) = 2d + 1 - 2i$. Таким образом, в силу (1) получим $\max\limits_{x \in A_1} D(x) = \max\{\lfloor (2j + \Delta)/2 \rfloor, \lfloor (2d + 1 - 2i + j + \Delta/2)/2 \rfloor\} = \max\limits_{\lfloor d/2 \rfloor < i < d - p} \{2d + 1 - 2i\} \leqslant d$.

Для $A_2$ имеем $D(x_l) = D(x_r) = j$. Из вершин $x_l$ и $x_r$ навстречу идут волны $\Delta$-интервалов, увеличивающих на 2 значения $D(x)$. Волны заканчиваются, когда $D(x)$ достигает значений $2\lceil d/2 \rceil - 1$. Учитывая $|A_2|$ и применяя (1), получим $\max\limits_{x \in A_2} D(x) = \lfloor (2j + 2(d - p - j) + \Delta/2 + 1)/2 \rfloor = d$.

Для $A_3$ имеем $D(x_r) = j$, $D(ir + r) = 2d - 1 - 2i$. На концах остальных $\Delta$-интервалов, входящих в $A_3$, $D(x) = j - 2$. Таким образом, сравнивая с множеством $A_1$ и применяя (1), получим $\max\limits_{x \in A_3} D(x) < \max\limits_{x \in A_1} D(x) < d$.

6) Пусть $x \in V_5$, где $1 \leqslant p \leqslant d - 1$.

Множество $V_5$ состоит из $r_i$-интервала, где $i = d - p$. При расчёте $D(x)$ для вершин $x \in [F, R]$ также учитываем $j = 1$. Используя (5), определяем значения $D(x)$ в вершинах $F$ и $R$. Делим вершины $r_i$ следующим образом: $A_1 = [F, F + \Delta/2]$, $A_2 = [F + \Delta/2, R - \Delta/2]$, $A_3 = [R - \Delta/2, R]$, $|A_1| = |A_3| = \Delta/2$, $|A_2| = (d - p - 1)\Delta + \Delta/2 + 1$.

Для $A_1$ и $A_3$ различаем два случая:

*а)* $D(F) = D(R) = 2p + 1$. Согласно (1), получаем $\max\limits_{x \in A_1} D(x) = \max\limits_{x \in A_3} D(x) = \lfloor (2p + 1 + 1 + \Delta/2)/2 \rfloor = 2p + 1 \leqslant d$;

*б)* $D(F) = D(R) = 2(d - p)$. Согласно (1), получаем $\max\limits_{x \in A_1} D(x) = \max\limits_{x \in A_3} D(x) = \lfloor (2(d - p) + 1 + \Delta/2)/2 \rfloor = d$.

Случай, когда $x \in A_2$, сводится к случаю 5, когда $x \in A_2$ и $j = 1$. ■

Из доказательства теоремы 1 следует наличие общей схемы структуры рассмотренных графов, что, вероятно, даст возможность разработки для них общего вида функции расстояний $D(x)$, зависящей от $d$ и параметра $p$.

## 2. Способы построения серии семейств циркулянтных сетей степени шесть

Можно выделить два способа получения серии циркулянтных сетей степени шесть.

П е р в ы й  с п о с о б. Пусть параметр $p$ последовательно пробегает значения на всем диапазоне (2) для каждого целого $d > 1$. Тогда получаем бесконечное множество $\Psi$ кольцевых циркулянтных сетей степени шесть и диаметров $d = 2, 3, \ldots$:

$$\Psi = \bigcup_{p=1,2,\ldots,d-1} \bigcup_{d>1} C(N; 1, s_2, s_3),$$

где $N$, $s_2$ и $s_3$ определяются формулами (3). Имеет место следующее свойство порядков графов полученной серии семейств.

**Лемма 2.** Число вершин $N$ графов вида (3) при всех $d > 1$ и $p = 1, 2, \ldots, d-1$ есть произведение двух взаимно простых нечётных чисел.

***Доказательство.*** Рассмотрим циркулянтный граф $C(N; 1, s_2, s_3)$ вида (3). Здесь $N = qr$, где $q = 2d - 2p + 1$, $r = 4pd - 4p^2 + 2p + 1$. Отсюда следует $r = 2pq + 1$, то есть $q$ и $r$ — взаимно простые числа при всех $d > 1$ и $p = 1, 2, \ldots, d-1$. ■

В табл. 3 дан пример представления значений порядков $N$ графов в виде произведений двух взаимно простых чисел для диаметров $2 \leqslant d \leqslant 8$ и $1 \leqslant p \leqslant d-1$.

Т а б л и ц а  3

**Представление порядков $N$ графов множества $\Psi$ в виде произведения двух взаимно простых чисел**

| $d$ | $p$ | | | | | | |
|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| 2 | $N = 3 \times 7$ | | | | | | |
| 3 | $N = 5 \times 11$ | $3 \times 13$ | | | | | |
| 4 | $N = 7 \times 15$ | $5 \times 21$ | $3 \times 19$ | | | | |
| 5 | $N = 9 \times 19$ | $7 \times 29$ | $5 \times 31$ | $3 \times 25$ | | | |
| 6 | $N = 11 \times 23$ | $9 \times 37$ | $7 \times 43$ | $5 \times 41$ | $3 \times 31$ | | |
| 7 | $N = 13 \times 27$ | $11 \times 45$ | $9 \times 55$ | $7 \times 57$ | $5 \times 51$ | $3 \times 37$ | |
| 8 | $N = 15 \times 31$ | $13 \times 53$ | $11 \times 67$ | $9 \times 73$ | $7 \times 71$ | $5 \times 61$ | $3 \times 43$ |

С помощью системы Wolfram Mathematica 10 был получен фрагмент одного из возможных построений семейств циркулянтных сетей из множества $\Psi$. Диаметр графов изменялся от $d = 2$ до 25, а параметр $p$ — от $p = 1$ до $p = d - 1$. В табл. 4 приведены описания найденных трёхмерных циркулянтных графов вида $C(N; 1, s_2, s_3)$: диаметры графов $3 \leqslant d \leqslant 10$, соответствующие им значения $1 \leqslant p \leqslant d-1$, порядки графов $N$ и образующие $s_2$ и $s_3$. На рис. 1 показан график зависимости $N$ от $p$ и $d$ для полученного фрагмента циркулянтных графов из множества $\Psi$.

Решим теперь задачу оптимизации для циркулянтных графов из множества $\Psi$: на множестве графов $\Psi$ заданного диаметра $d > 1$ найти функцию $p = p(d)$, которая задаёт максимум функции $N = N(p)$ при всех $d > 1$.

Таблица 4

**Параметры описания графов множества $\Psi$ при $3 \leqslant d \leqslant 10$**

| $d$ | $p$ | $N$ | $s_2$ | $s_3$ | $d$ | $p$ | $N$ | $s_2$ | $s_3$ | $d$ | $p$ | $N$ | $s_2$ | $s_3$ | $d$ | $p$ | $N$ | $s_2$ | $s_3$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 3 | 1 | 55 | 20 | 24 | 6 | 3 | 301 | 123 | 135 | 8 | 3 | 737 | 329 | 341 | 9 | 7 | 355 | 128 | 156 |
| 3 | 2 | 39 | 9 | 17 | 6 | 4 | 205 | 74 | 90 | 8 | 4 | 657 | 284 | 300 | 9 | 8 | 147 | 33 | 65 |
| 4 | 1 | 105 | 43 | 47 | 6 | 5 | 93 | 21 | 41 | 8 | 5 | 497 | 203 | 223 | 10 | 1 | 741 | 349 | 353 |
| 4 | 2 | 105 | 38 | 46 | 7 | 1 | 351 | 160 | 164 | 8 | 6 | 305 | 110 | 134 | 10 | 2 | 1173 | 548 | 556 |
| 4 | 3 | 57 | 13 | 25 | 7 | 2 | 495 | 221 | 229 | 8 | 7 | 129 | 29 | 57 | 10 | 3 | 1365 | 631 | 643 |
| 5 | 1 | 171 | 74 | 78 | 7 | 3 | 495 | 214 | 226 | 9 | 1 | 595 | 278 | 282 | 10 | 4 | 1365 | 622 | 638 |
| 5 | 2 | 203 | 83 | 91 | 7 | 4 | 399 | 163 | 179 | 9 | 2 | 915 | 423 | 431 | 10 | 5 | 1221 | 545 | 565 |
| 5 | 3 | 155 | 56 | 68 | 7 | 5 | 255 | 92 | 112 | 9 | 3 | 1027 | 468 | 480 | 10 | 6 | 981 | 424 | 448 |
| 5 | 4 | 75 | 17 | 33 | 7 | 6 | 111 | 25 | 49 | 9 | 4 | 979 | 437 | 453 | 10 | 7 | 693 | 283 | 311 |
| 6 | 1 | 253 | 113 | 117 | 8 | 1 | 465 | 215 | 219 | 9 | 5 | 819 | 354 | 374 | 10 | 8 | 405 | 146 | 178 |
| 6 | 2 | 333 | 144 | 152 | 8 | 2 | 689 | 314 | 322 | 9 | 6 | 595 | 243 | 267 | 10 | 9 | 165 | 37 | 73 |

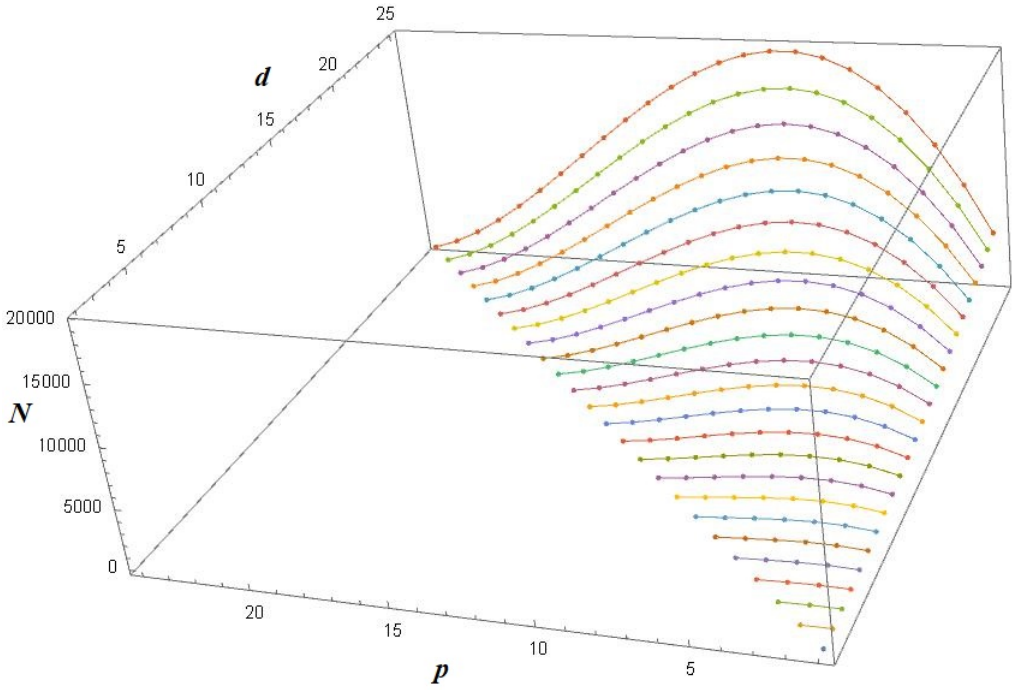



Рис. 1. График зависимости порядка $N$ циркулянтов из $\Psi$ для $d = 2, \ldots, 25$ и $p = 1, \ldots, d-1$

**Теорема 2.** Для любого целого $d > 1$ максимум $N = N(p)$, определяемого формулами (3), достигается при

$$p(d) = p^* = \begin{cases} \lfloor d/3 \rfloor, & \text{если } d \equiv 0 \pmod 3 \text{ или } d \equiv 1 \pmod 3, \\ \lceil d/3 \rceil, & \text{если } d \equiv 2 \pmod 3 \text{ или } d \equiv 1 \pmod 3. \end{cases} \tag{9}$$

***Доказательство.*** Рассмотрим циркулянтный граф $C(N; 1, s_2, s_3)$ вида (3). Функция $N$ — кубический полином относительно $p$ для любого заданного $d$. Надо найти такую целочисленную функцию $p(d)$, при которой значение $N$ равно максимуму для любого $d > 1$. Для этого вычислим производную $N$ по $p$ и приравняем её нулю: $\dfrac{dN}{dp} = 24p^2 - 16(2d+1)p + 8d(d+1) = 0$. Полученное квадратное уравнение относительно $p$ имеет коэффициенты $a = 24$, $b = -(32d + 16)$, $c = 8d^2 + 8d$. Дискриминант $\delta = b^2 - 4ac = 16^2(d^2 + d + 1) > 0$. Следовательно, $N$ имеет один локальный максимум,

когда $p_1 = (2d + 1 - \sqrt{d^2 + d + 1})/3$ (второе решение $p_2 = (2d + 1 + \sqrt{d^2 + d + 1})/3 \geqslant d$ не подходит). Так как $d < \sqrt{d^2 + d + 1} < d + 1$ и соответственно $d/3 < p_1 < (d + 1)/3$, взяв ближайшее целое, получим для любого $d > 1$ значения $p(d)$, равные (9). Подставляя найденные значения $p$ в (3), получим (10) (см. далее), а также соответствующие значения образующих максимального графа. ■

В т о р о й  с п о с о б. Если в качестве $p$ взять любую целочисленную функцию от $d$, удовлетворяющую условию $1 \leqslant p(d) < d$, то можно синтезировать новые бесконечные семейства циркулянтных сетей. Ниже представлены два примера полученных таким способом семейств циркулянтных сетей степени шесть, принадлежащих $\Psi$.

**Пример 1.** Пусть $p(d) = \lceil d/2 \rceil$, где $d > 1$. Тогда

$$C(N; 1, s_2, s_3) = \begin{cases} C(d^3 + 2d^2 + 2d + 1; 1, (d^3 + d^2 - d)/2, (d^3 + d^2 + 3d)/2) \text{ при чётных } d, \\ C(d^3 + d^2 + d; 1, (d^3 - 3)/2 - d, (d^3 - 3)/2 + d + 2) \text{ при нечётных } d. \end{cases}$$

Новое семейство из примера 1 по соотношению $N/d$ лучше семейств, найденных в [9, 10, 14, 15].

**Пример 2.** Пусть $p(d) = p^*$, где $p^*$ определяется соотношением (9). Тогда семейство циркулянтных графов $C(N; 1, s_2, s_3)$ диаметра $d > 1$ с максимальным $N$ и образующими, представленными в виде полиномов третьей степени от $d$, описывается следующим образом:

$$N(d) = \begin{cases} \dfrac{32}{27}d^3 + \dfrac{16}{9}d^2 + 2d + 1, & \text{если } d \equiv 0 \pmod 3, \\ 32\lfloor d/3 \rfloor^3 + 48\lfloor d/3 \rfloor^2 + 22\lfloor d/3 \rfloor + 3, & \text{если } d \equiv 1 \pmod 3, \\ 32\lfloor d/3 \rfloor^3 + 80\lfloor d/3 \rfloor^2 + 70\lfloor d/3 \rfloor + 21, & \text{если } d \equiv 2 \pmod 3), \end{cases} \tag{10}$$

$$(s_2(d), s_3(d)) = \begin{cases} \left(\dfrac{16}{27}d^3 + \dfrac{4}{9}d^2,\ s_2 + \dfrac{4}{3}d\right), & \text{если } d \equiv 0 \pmod 3, \\ \left(\dfrac{16}{27}d^3 + \dfrac{4}{9}d^2 - \dfrac{2}{3}d + \dfrac{17}{27},\ s_2 + \dfrac{4}{3}d - \dfrac{4}{3}\right) \text{ или} & \\ \left(\dfrac{16}{27}d^3 + \dfrac{4}{9}d^2 - \dfrac{4}{3}d - \dfrac{46}{27},\ s_2 + \dfrac{4}{3}d + \dfrac{8}{3}\right), & \text{если } d \equiv 1 \pmod 3, \\ \left(\dfrac{16}{27}d^3 + \dfrac{4}{9}d^2 - \dfrac{2}{9}d - \dfrac{29}{27},\ s_2 + \dfrac{4}{3}d + \dfrac{4}{3}\right), & \text{если } d \equiv 2 \pmod 3. \end{cases}$$

Семейство из примера 2 по соотношению $N/d$ превосходит семейства, полученные в [9–11, 14,15]. Для всех диаметров $d \equiv 0 \pmod 3$ и $d \equiv 2 \pmod 3$ максимальный порядок $N(d)$, равный (10), совпадает с максимумом $N$, найденным в [16], а при $d \equiv 1 \pmod 3$ оказывается меньше на величину $4(2\lfloor d/3 \rfloor + 1)$. Отметим, что при $d \equiv 1 \pmod 3$ существуют два набора образующих третьей степени от $d$, которые задают максимум $N(d)$, равный (10).

## Заключение

Получена серия параметрически описываемых бесконечных семейств кольцевых циркулянтных сетей степени шесть, включающая графы максимального порядка для заданного диаметра. Это является новым результатом в теории циркулянтных сетей, дающим возможность синтеза ранее неизвестных семейств с меняющимся диаметром, а также при фиксированном диаметре $d > 1$ построения серии из $d - 1$ графов. Ранее были известны только отдельные бесконечные семейства циркулянтов. Другая

особенность полученного результата — наличие общей схемы структуры графов получающихся семейств — даёт возможность разработки для них общих аналитических методов поиска кратчайших путей, что подтверждено на примере семейства из [17], являющегося частным случаем параметрически описываемых бесконечных семейств. Получение новых серий семейств сетей, построенных на других типах определяющих соотношений между порядком и образующими графа, и эффективных аналитических алгоритмов парной маршрутизации для них является одним из направлений будущей работы и представляет интерес с практической точки зрения, так как циркулянтные графы степени шесть известны как одна из перспективных топологий для сетей на кристалле.

Автор выражает благодарность О. Г. Монахову за экспериментальные результаты, проведённые с помощью системы Wolfram Mathematica 10.

## ЛИТЕРАТУРА

1. *Монахова Э. А.* Структурные и коммуникативные свойства циркулянтных сетей // Прикладная дискретная математика. 2011. № 3. С. 92–115.
2. *Monakhova E. A.* A survey on undirected circulant graphs // Discrete Math. Algorithms Appl. 2012. No. 4. `https://www.researchgate.net/publication/267143246_A_survey_on_undirected_circulant_graphs`.
3. *Perez-Roses H.* Algebraic and computer-based methods in the undirected degree/diameter problem — A brief survey // Electr. J. Graph Theory Appl. 2014. No. 2(2). P. 166–190.
4. *Bermond J.-C., Comellas F., and Hsu D. F.* Distributed loop computer networks: a survey // J. Parallel Distributed Comput. 1995. No. 24. P. 2–10.
5. *Hwang F. K.* A survey on multi-loop networks // Theor. Comput. Sci. 2003. No. 299. P. 107–121.
6. *Romanov A., Amerikanov A., and Lezhnev E.* Analysis of approaches for synthesis of networks-on-chip by using circulant topologies // J. Physics: Conf. Ser. 2018. V. 1050. P. 1–12.
7. *Romanov A. Yu.* Development of routing algorithms in networks-on-chip based on ring circulant topologies // Heliyon. 2019. V. 5. No. 4. P. 1–23.
8. *Романов А. Ю., Ведмидь Е. А., Монахова Э. А.* Проектирование сетей на кристалле с топологией кольцевой циркулянт с тремя образующими: разработка алгоритмов маршрутизации // Информационные технологии. 2019. № 25(9). С. 522-530.
9. *Yebra J. L. A., Fiol M. A., Morillo P., and Alegre I.* The diameter of undirected graphs associated to plane tessellations // Ars Combinatoria. 1985. No. 20B. P. 159–172.
10. *Wong C. K. and Coppersmith D.* A combinatorial problem related to multimodule memory organizations // J. Assoc. Comput. Mach. 1974. No. 21. P. 392–402.
11. *Chen S. and Jia X.-D.* Undirected loop networks // Networks. 1993. No. 23. P. 257–260.
12. *Barriere L., Fabrega J., Simo E., and Zaragoza M.* Fault-tolerant routings in chordal ring networks // Networks. 2000. V. 36(3). P. 180–190.
13. *Thomson A. and Zhou S.* Gossiping and routing in undirected triple-loop networks // Networks. 2010. No. 55(4). P. 341–349.
14. *Liestman A. L., Opatrny J., and Zaragoza M.* Network properties of double and triple fixed-step graphs // Int. J. Found. Comp. Sci. 1998. V. 9. P. 57–76.
15. *Jha P. K.* A family of efficient six-regular circulants representable as a Kronecker product // Discr. Appl. Math. 2016. V. 203. P. 72–84.
16. *Monakhova E.* Optimal triple loop networks with given transmission delay: Topological design and routing // Intern. Network Optimization Conf. (INOC'2003), Evry/Paris, France, 2003. P. 410–415.

17. *Монахова Э. А., Монахов О. Г.* Динамический алгоритм парной маршрутизации для аналитически задаваемых семейств циркулянтных сетей степени шесть // Сб. статей XIX Междунар. науч.-технич. конф. «Проблемы информатики в образовании, управлении, экономике и технике». Пенза: ПДЗ, 2019. С. 30–37.

REFERENCES

1. *Monakhova E. A.* Strukturnye i kommunikativnye svoystva tsirkulyantnykh setey [Structural and communicative properties of circulant networks]. Prikladnaya Diskretnaya Matematika, 2011, no. 3, pp. 92–115. (in Russian)
2. *Monakhova E. A.* A survey on undirected circulant graphs. Discrete Math. Algorithms Appl., 2012, no. 4. `https://www.researchgate.net/publication/267143246_A_survey_on_undirected_circulant_graphs`.
3. *Perez-Roses H.* Algebraic and computer-based methods in the undirected degree/diameter problem — A brief survey. Electr. J. Graph Theory Appl., 2014, no. 2(2), pp. 166–190.
4. *Bermond J.-C., Comellas F., and Hsu D. F.* Distributed loop computer networks: A survey. J. Parallel Distributed Comput., 1995, no. 24, pp. 2–10.
5. *Hwang F. K.* A survey on multi-loop networks. Theor. Comput. Sci., 2003, no. 299, pp. 107–121.
6. *Romanov A., Amerikanov A., and Lezhnev E.* Analysis of approaches for synthesis of networks-on-chip by using circulant topologies. J. Physics: Conf. Ser., 2018, vol. 1050, pp. 1–12.
7. *Romanov A. Yu.* Development of routing algorithms in networks-on-chip based on ring circulant topologies. Heliyon, 2019, vol. 5, no. 4, pp. 1–23.
8. *Romanov A. Yu., Vedmid E. A., and Monakhova E. A.* Proektirovanie setej na kristalle s topologiej kol'cevoj cirkulyant s tremya obrazuyushchimi: razrabotka algoritmov marshrutizacii [Designing networks-on-chip based on triple loop (circulant) networks: routing algorithm development]. Informacionnye Tekhnologii, 2019, no. 25(9), pp. 522–530. (in Russian)
9. *Yebra J. L. A., Fiol M. A., Morillo P., and Alegre I.* The diameter of undirected graphs associated to plane tessellations. Ars Combinatoria, 1985, no. 20B, pp. 159–172.
10. *Wong C. K. and Coppersmith D.* A combinatorial problem related to multimodule memory organizations. J. Assoc. Comput. Mach., 1974, no. 21, pp. 392–402.
11. *Chen S. and Jia X.-D.* Undirected loop networks. Networks, 1993, no. 23, pp. 257–260.
12. *Barriere L., Fabrega J., Simo E., and Zaragoza M.* Fault-tolerant routings in chordal ring networks. Networks, 2000, no. 36(3), pp. 180–190.
13. *Thomson A. and Zhou S.* Gossiping and routing in undirected triple-loop networks. Networks, 2010, no. 55(4), pp. 341–349.
14. *Liestman A. L., Opatrny J., and Zaragoza M.* Network properties of double and triple fixed-step graphs. Int. J. Found. Comp. Sci., 1998, vol. 9, pp. 57–76.
15. *Jha P. K.* A family of efficient six-regular circulants representable as a Kronecker product. Discr. Appl. Math., 2016, vol. 203, pp. 72–84.
16. *Monakhova E.* Optimal triple loop networks with given transmission delay: Topological design and routing. Intern. Network Optimization Conf. (INOC'2003), Evry/Paris, France, 2003, pp. 410–415.
17. *Monakhova E. A. and Monakhov O. G.* Dinamicheskij algoritm parnoj marshrutizacii dlya analiticheski zadavaemyh semejstv cirkulyantnyh setej stepeni shest' [A dynamic algorithm of two-terminal routing for analytically described families of degree six circulant networks]. Proc. XIX Intern. Conf. "Problemy Informatiki v Obrazovanii, Upravlenii, Ekonomike i Tekhnike", Penza, PDZ Publ., 2019, pp. 30–37. (in Russian)

# МАТЕМАТИЧЕСКИЕ ОСНОВЫ ИНФОРМАТИКИ И ПРОГРАММИРОВАНИЯ

УДК 510.52

## О ГЕНЕРИЧЕСКОЙ СЛОЖНОСТИ ЭКЗИСТЕНЦИАЛЬНЫХ ТЕОРИЙ[1]

А. Н. Рыбалов

*Институт математики им. С. Л. Соболева СО РАН, г. Омск, Россия*

Многие задачи о конечных графах и конечных полях могут быть сформулированы на языке универсальной алгебраической геометрии, в рамках которой эти объекты рассматриваются как алгебраические системы в заданном языке. Алгебраическая геометрия над этими объектами тесным образом связана со свойствами экзистенциальных теорий. С практической точки зрения важнейшими являются вопросы разрешимости и вычислительной сложности этих теорий. В данной работе изучается вычислительная сложность экзистенциальных теорий алгебраических систем конечного предикатного языка с равенством. Известно, что для любой алгебраической системы с более чем одноэлементным основным множеством эта теория является NP-трудной (NP-полной, если основное множество конечно). Поэтому возникает вопрос о генерической сложности экзистенциальной теории алгебраической системы конечного предикатного языка. Доказывается, что при условиях $\mathrm{P} \neq \mathrm{NP}$ и $\mathrm{P} = \mathrm{BPP}$ для распознавания этой теории не существует полиномиального сильно генерического алгоритма.

**Ключевые слова:** *генерическая сложность, конечная алгебраическая система, экзистенциальная теория.*

DOI 10.17223/20710410/49/9

## ON GENERIC COMPLEXITY OF THE EXISTENTIAL THEORIES

A. N. Rybalov

*Sobolev Institute of Mathematics, Omsk, Russia*

**E-mail:** alexander.rybalov@gmail.com

Many problems about finite graphs and finite fields can be formulated in the universal algebraic geometry, where these objects are considered as algebraic structures in the given language. Algebraic geometry over such objects is closely related to properties of existential theories. From a practical point of view, the most important are questions about decidability and computational complexity of these theories. In this paper we study the computational complexity of existential theories of algebraic structures of finite predicate language. It is known that the existential theory of any algebraic structure with more than one element is NP-hard. We prove that under the conditions $\mathrm{P} \neq \mathrm{NP}$ and $\mathrm{P} = \mathrm{BPP}$, for this theories there is no polynomial strongly generic

[1]Работа поддержана грантом РНФ № 19-11-00209.

algorithm. To prove this theorem we use the method of generic amplification, which allows to construct generically undecidable problems from the problems undecidable in the classical sense. The main ingredient of this method is a technique of cloning, which unites inputs of the problem together in the large enough sets of equivalent inputs. Equivalence is understood in the sense that the problem is solved similarly for them.

**Keywords:** *generic complexity, finite algebraic structure, existential theory.*

## Введение

В XX веке, в связи с бурным развитием компьютерной техники и прикладной математики, на первый план вышли исследования различных конечных комбинаторных и алгебраических объектов. Например, конечные графы находят многочисленные приложения при решении практических задач, связанных с сетями, маршрутами, классификацией объектов и т. д. Другой пример — это конечные поля, без которых немыслимы современная криптография и теория помехоустойчивой передачи информации. Классическими подходами к изучению конечных объектов являются алгебраический и комбинаторный. Новый подход — логический и теоретико-модельный — родился внутри так называемой универсальной алгебраической геометрии [1]. В рамках этого подхода изучаемые объекты рассматриваются как алгебраические системы в заданном языке (сигнатуре). Многие практически важные задачи о конечных объектах можно формулировать как задачи, связанные с решением систем уравнений над соответствующими алгебраическими системами, что приводит к необходимости развития алгебраической геометрии. Алгебраическая геометрия тесным образом связана со свойствами экзистенциальных теорий. С практической точки зрения важнейшими являются вопросы разрешимости и вычислительной сложности этих теорий.

Генерический подход к алгоритмическим проблемам предложен в [2]. В рамках этого подхода алгоритмическая проблема рассматривается не на всём множестве входов, а на некотором подмножестве «почти всех» входов. Такие входы образуют так называемое генерическое множество. Понятие «почти все» формализуется введением естественной меры на множестве входных данных. С точки зрения практики алгоритмы, решающие быстро проблему на генерическом множестве, так же хороши, как и быстрые алгоритмы для всех входов. Классическим примером такого алгоритма является симплекс-метод — он за полиномиальное время решает задачу линейного программирования для большинства входных данных, но имеет экспоненциальную сложность в худшем случае. Более того, может так оказаться, что проблема трудноразрешима или вообще неразрешима в классическом смысле, но легкоразрешима на генерическом множестве.

В данной работе изучается генерическая сложность экзистенциальных теорий алгебраических систем конечного предикатного языка с равенством. Для любой алгебраической системы с более чем одноэлементным основным множеством к этой теории может быть полиномиально сведена проблема выполнимости булевых формул [3]. Таким образом, проблема распознавания любой такой теории является NP-трудной (NP-полной, если основное множество конечно) и, при условии $\mathrm{P} \neq \mathrm{NP}$, не существует полиномиального алгоритма, распознающего её для всех входов. В работе доказывается, что при условиях $\mathrm{P} \neq \mathrm{NP}$ и $\mathrm{P} = \mathrm{BPP}$ для распознавания этой теории не существует полиномиального сильно генерического алгоритма. Класс BPP состоит из проблем, разрешимых за полиномиальное время на вероятностных машинах Тьюрин-

га. В [4] доказано, что равенство P = BPP следует из весьма правдоподобных гипотез о вычислительной сложности некоторых трудных проблем.

## 1. Предварительные сведения

Напомним некоторые определения из математической логики и теории моделей [1]. *Сигнатурой* называется множество $\sigma$, состоящее из предикатных, функциональных и константных символов. *Алгебраическая система* сигнатуры $\sigma$ есть набор $\mathfrak{A} = \langle A, \sigma \rangle$, где $A$ — непустое основное множество, причём каждому предикатному символу сигнатуры $\sigma$ сопоставлен некоторый предикат на множестве $A$, каждому функциональному символу из $\sigma$ — функция на множестве $A$ со значениями в множестве $A$, а каждому константному символу из $\sigma$ — элемент из $A$.

Формула логики первого порядка сигнатуры $\sigma$, в которой каждая переменная связана некоторым квантором, называется *предложением*. Тот факт, что предложение $\Phi$ сигнатуры $\sigma$ истинно в алгебраической системе $\mathfrak{A} = \langle A, \sigma \rangle$, обозначается как $\mathfrak{A} \models \Phi$. Предложение $\Phi$ сигнатуры $\sigma$ называется *экзистенциальным* (или $\exists$-предложением), если оно имеет вид

$$\Phi = \exists x_1 \ldots \exists x_n \varphi(x_1, \ldots, x_n),$$

где $\varphi$ — бескванторная формула сигнатуры $\sigma$. *Элементарной теорией* алгебраической системы $\mathfrak{A}$ называется множество $Th(\mathfrak{A})$ всех предложений сигнатуры $\sigma$, истинных в $\mathfrak{A}$. Множество всех $\exists$-предложений теории $Th(\mathfrak{A})$ называется *экзистенциальной* теорией $Th_{\exists}(\mathfrak{A})$ алгебраической системы $\mathfrak{A}$.

Будем использовать числа Каталана $C_n$, которые определяются следующим образом:

$$C_n = \frac{1}{n+1}\binom{2n}{n}.$$

Здесь $\binom{2n}{n}$ — соответствующий биномиальный коэффициент. В дальнейшем понадобится следующее утверждение о числах Каталана.

**Лемма 1.** Для $n > m$ имеет место

$$\frac{C_{n-m}}{C_n} > \frac{1}{4^m}.$$

***Доказательство.*** Оценим отношение чисел Каталана:

$$\frac{C_{n-m}}{C_n} = \frac{n+1}{n-m+1}\,\frac{\binom{2(n-m)}{n-m}}{\binom{2n}{n}} = \frac{n+1}{n-m+1}\,\frac{\dfrac{(2(n-m))!}{(n-m)!(n-m)!}}{\dfrac{(2n)!}{n!n!}} >$$

$$> \frac{(2(n-m))!n!n!}{(n-m)!(n-m)!(2n)!} = \frac{n!}{(n-m)!}\,\frac{2(n-m)\ldots(n-m+1)}{2n\ldots(n+1)} =$$

$$= \frac{n(n-1)\ldots(n-m+1)2(n-m)\ldots(n-m+1)}{2n\ldots(n+1)} =$$

$$= \frac{(n\ldots(n-m+1))^2}{2n\ldots(2(n-m)+1)} > \Big(\frac{(n-1)\ldots(n-m)}{2(n-1)\ldots(2n-2m)}\Big)^2 > \frac{1}{2^{2m}} = \frac{1}{4^m}.$$

Лемма 1 доказана. ■

## 2. Генерические алгоритмы

Пусть $I$ — некоторое множество входов. Для подмножества $S \subseteq I$ определим последовательность

$$\rho_n(S) = \frac{|S \cap I_n|}{|I_n|},\ n = 1, 2, 3, \ldots,$$

где $I_n$ — множество входов размера $n$. Заметим, что $\rho_n(S)$ — это вероятность попасть в $S$ при случайной и равновероятной генерации входов из $I_n$. *Асимптотической плотностью* $S$ назовём предел

$$\rho(S) = \overline{\lim_{n\to\infty}} \rho_n(S).$$

Множество $S$ называется *генерическим*, если $\rho(S) = 1$, и *пренебрежимым*, если $\rho(S) = 0$. Следуя [2], назовём множество $S$ *сильно пренебрежимым*, если последовательность $\rho_n(S)$ экспоненциально быстро сходится к 0, т. е. существуют константы $\sigma$, $0 < \sigma < 1$, и $C > 0$, такие, что для любого $n$

$$\rho_n(S) < C\sigma^n.$$

Теперь $S$ называется *сильно генерическим*, если его дополнение $I \setminus S$ сильно пренебрежимо.

Алгоритм $\mathcal{A}$ с множеством входов $I$ и множеством выходов $J \cup \{?\}$ $(? \notin J)$ называется (*сильно*) *генерическим*, если

1) $\mathcal{A}$ останавливается на всех входах из $I$;
2) множество $\{x \in I : \mathcal{A}(x) \neq ?\}$ является (сильно) генерическим.

Генерический алгоритм $\mathcal{A}$ вычисляет функцию $f : I \to J$, если $(\mathcal{A}(x) = y \in J) \Rightarrow$ $\Rightarrow (f(x) = y)$ для всех $x \in I$. Ситуация $\mathcal{A}(x) = ?$ означает, что $\mathcal{A}$ не может вычислить функцию $f$ на аргументе $x$. Но условие 2 гарантирует, что $\mathcal{A}$ корректно вычисляет $f$ на почти всех входах (входах из генерического множества).

## 3. Представление экзистенциальных предложений

Рассмотрим естественное представление экзистенциальных предложений с помощью двоичных деревьев. Это представление, с одной стороны, настолько же компактно, как и стандартное представление строками символов (с точностью до линейного множителя). С другой стороны, оно удобно для различного рода подсчётов. Кроме того, достаточно просто написать компьютерную программу для случайной генерации предложений, заданных с помощью этого представления.

Зафиксируем конечную предикатную сигнатуру

$$\sigma = \{P_1^{(a_1)}, \ldots, P_k^{(a_k)}, c_1, \ldots, c_l\},$$

где $P_i$ — предикаты (включая двуместный предикат равенства); $c_i$ — константы. Положим

$$A = \max_{i=1,\ldots,k} \{a_i\}.$$

Пусть экзистенциальное предложение $\Phi$ сигнатуры $\sigma$ имеет вид

$$\Phi = \exists x_1 \ldots \exists x_t \phi,$$

где $\phi$ — бескванторная формула, полученная с помощью конъюнкций и дизъюнкций из атомарных формул вида $P_i(x_1, \ldots, x_{a_i})$ или их отрицаний. Структуру формулы $\phi$ естественно представлять в виде бинарного дерева $T_\phi$, такого, что внутренние вершины $T_\phi$

помечены символами $\vee$ и $\wedge$, а листья $T_\phi$ — простыми атомарными формулами или их отрицаниями. Если $T_\phi$ имеет $n$ листьев, то не более $An$ переменных могут встретиться в $T_\phi$, поэтому в дальнейшем будем полагать, что все переменные $T_\phi$ лежат в множестве $\{x_1, \ldots, x_{An}\}$. Под размером предложения $\Phi$ будем понимать число листьев в дереве $T_\phi$. Для упрощения подсчётов считается, что предложение $\Phi$ размера $n$ зависит от всех переменных $x_1, \ldots, x_{An}$ и все эти переменные связаны кванторами.

Обозначим через $\mathcal{F}$ множество экзистенциальных предложений, представленных описанным способом.

**Лемма 2.** Число экзистенциальных предложений размера $n$ есть

$$|\mathcal{F}_n| = 2^{n-1} C_{n-1} \Big( 2 \sum_{i=1}^{k} (An + l)^{a_i} \Big)^n. \tag{1}$$

***Доказательство.*** Любое предложение из $\mathcal{F}$ размера $n$ состоит из кванторной приставки и бинарного дерева с $n$ листьями и $n-1$ внутренними вершинами. Известно (см., например, [5]), что существует $C_{n-1}$ неразмеченных бинарных деревьев с $n$ листьями. Каждая внутренняя вершина такого дерева может быть помечена символами $\vee$ или $\wedge$, поэтому есть всего $2^{n-1}$ таких разметок. Каждый из $n$ листьев может быть помечен одним из $a_i$-местных предикатов $P_i$, $i = 1, \ldots, k$, либо его отрицанием, в которые можно подставлять на каждое из $a_i$ мест либо одну из переменных $x_j$, $j = 1, \ldots, An$, либо константу $c_j$, $j = 1, \ldots, l$. Таким образом, получаем формулу (1). ■

Для любого экзистенциального предложения $\Phi = \exists x_1 \ldots \exists x_t \phi$ рассмотрим множество предложений

$$eq(\Phi) = \left\{ \exists x_1 \ldots \exists x_{2n} \big( \phi \vee ((x_1 \neq x_1) \wedge \psi) \big) \right\},$$

где $n > t$; $\psi$ — произвольная бескванторная формула от переменных $x_1, \ldots, x_n$. Легко видеть, что все предложения из $eq(\Phi)$ эквивалентны $\Phi$ в том смысле, что они истинны или ложны одновременно с $\Phi$.

**Лемма 3.** Для любого экзистенциального предложения $\Phi$ имеет место

$$\frac{|eq(\Phi) \cap \mathcal{F}_n|}{|\mathcal{F}_n|} > \frac{1}{\big(16k(An + l)^A\big)^{m+2}}$$

для любого $n > m + 2$, где $m$ — размер предложения $\Phi$.

***Доказательство.*** Рассмотрим любое предложение из множества $eq(\Phi)$ размера $n > m + 2$. Заметим, что число вершин в дереве $T_\psi$ из этого предложения равно $n - m - 2$. Теперь аналогично тому, как это делалось в доказательстве леммы 2, можно подсчитать значение

$$|eq(\Phi) \cap \mathcal{F}_n| = 2^{n-m-3} C_{n-m-3} \Big( 2 \sum_{i=1}^{k} (An + l)^{a_i} \Big)^{n-m-2}.$$

В результате получим

$$\frac{|eq(\Phi) \cap \mathcal{F}_n|}{|\mathcal{F}_n|} = \frac{2^{n-m-3} C_{n-m-3} \Big( 2 \sum_{i=1}^{k} (An + l)^{a_i} \Big)^{n-m-2}}{2^{n-1} C_{n-1} \Big( 2 \sum_{i=1}^{k} (An + l)^{a_i} \Big)^{n}} =$$

$$= \frac{1}{2^{m+2}\Big(2\sum\limits_{i=1}^{k}(An+l)^{a_i}\Big)^{m+2}}\ \frac{C_{n-m-3}}{C_{n-1}} >$$

$$> \frac{1}{2^{m+2}\Big(2\sum\limits_{i=1}^{k}(An+l)^{a_i}\Big)^{m+2}}\ \frac{1}{4^{m+2}} > \frac{1}{\big(16k(An+l)^{A}\big)^{m+2}}.$$

Здесь использована лемма 1 для оценки отношения чисел Каталана. ■

## 4. Основной результат

**Теорема 1.** Пусть $\mathfrak{A}$ — алгебраическая система конечной предикатной сигнатуры с более чем одним элементом в основном множестве. Если существует сильно генерический полиномиальный алгоритм, распознающий $Th_{\exists}(\mathfrak{A})$, то существует вероятностный полиномиальный алгоритм, распознающий $Th_{\exists}(\mathfrak{A})$ на всём множестве предложений.

***Доказательство.*** Допустим, что существует сильно генерический полиномиальный алгоритм $\mathcal{A}$, распознающий $Th_{\exists}(\mathfrak{A})$. Тогда множество

$$G(\mathcal{A}) = \{\Phi \in \mathcal{F} : \mathcal{A}(\Phi) \neq ?\}$$

является сильно генерическим. Построим вероятностный полиномиальный алгоритм $\mathcal{B}$, определяющий истинность любого предложения $\Phi$. На предложении $\Phi$ размера $n$ алгоритм $\mathcal{B}$ работает следующим образом:

1) Генерирует случайное предложение $\Psi$ из множества $eq(\Phi)$ размера $n^2$.
2) Вычисляет $\mathcal{A}(\Psi)$.
3) Если $\mathcal{A}(\Psi) \neq ?$, то определяет истинность $\Phi$.
4) Если $\mathcal{A}(\Psi) = ?$, то выдаёт ответ «НЕТ».

Заметим, что алгоритм выдает правильный ответ на шаге 3, а на шаге 4 может выдать неправильный ответ. Нужно доказать, что вероятность того, что ответ выдаётся на шаге 4, меньше $1/2$.

Вероятность того, что случайное предложение вида $\Psi$ из $eq(\Phi)_{n^2}$ не попадёт в $G(\mathcal{A})$, не больше

$$\frac{|(\mathcal{F} \setminus G(\mathcal{A}))_{n^2}|}{|eq(\Phi)_{n^2}|} = \frac{|(\mathcal{F} \setminus G(\mathcal{A}))_{n^2}|}{|\mathcal{F}_{n^2}|}\ \frac{|\mathcal{F}_{n^2}|}{|eq(\Phi)_{n^2}|}.$$

Так как $G(\mathcal{A})$ сильно генерическое, то существует константа $\alpha > 0$, такая, что

$$\frac{|(\mathcal{F} \setminus G(\mathcal{A}))_{n^2}|}{|\mathcal{F}_{n^2}|} < \frac{1}{2^{\alpha n^2}}$$

для любого $n$. С другой стороны, по лемме 3

$$\frac{|\mathcal{F}_{n^2}|}{|eq(\Phi)_{n^2}|} < \big(16k(An^2+l)^{A}\big)^{n+2}.$$

Поэтому искомая вероятность не больше

$$\frac{\big(16k(An^2+l)^{A}\big)^{n+2}}{2^{\alpha n^2}} = \frac{2^{(n+2)\log\left(16k(An^2+l)^{A}\right)}}{2^{\alpha n^2}}$$

и при больших $n$ меньше $1/2$. Это означает, что вероятность выдачи ответа на шаге 4 меньше $1/2$.

Полиномиальность описанного алгоритма следует из существования процедуры генерации двоичного дерева размера $N$ за время, полиномиально ограниченное от $N$. Эта процедура описана в [6]. ■

Непосредственно из теоремы 1 следует

**Теорема 2.** Пусть $\mathfrak{A}$ — алгебраическая система конечной предикатной сигнатуры с более чем одним элементом в основном множестве. Если $\mathrm{P} \neq \mathrm{NP}$ и $\mathrm{P} = \mathrm{BPP}$, то не существует сильно генерического полиномиального алгоритма, распознающего $Th_{\exists}(\mathfrak{A})$.

Автор выражает благодарность рецензенту за полезные замечания и предложения по улучшению текста статьи.

## ЛИТЕРАТУРА

1. *Даниярова Э.Ю., Мясников А.Г., Ремесленников В.Н.* Алгебраическая геометрия над алгебраическими системами. Новосибирск: СО РАН, 2016. 288 с.
2. *Kapovich I., Miasnikov A., Schupp P., and Shpilrain V.* Generic-case complexity, decision problems in group theory and random walks // J. Algebra. 2003. V. 264. No. 2. P. 665–694.
3. *Гэри М., Джонсон Д.* Вычислительные машины и труднорешаемые задачи. М.: Мир, 1982. 419 с.
4. *Impagliazzo R. and Wigderson A.* P = BPP unless E has subexponential circuits: Derandomizing the XOR Lemma // Proc. 29th STOC. El Paso: ACM, 1997. P. 220–229.
5. *Кнут Д.* Искусство программирования. М.: Вильямс, 2010. 720 с.
6. *Рыбалов А. Н.* О генерической сложности проблемы общезначимости булевых формул // Прикладная дискретная математика. 2016. № 2(32). С. 119–126.

## REFERENCES

1. *Daniyarova E. Y., Myasnikov A. G., and Remeslennikov V. N.* Algebraicheskaya geometriya nad algebraicheskimi sistemami [Algebraic Geometry over Algebraic Structures]. Novosibirsk, SB RAS Publ., 2016. 288 p (in Russian).
2. *Kapovich I., Miasnikov A., Schupp P., and Shpilrain V.* Generic-case complexity, decision problems in group theory and random walks. J. Algebra, 2003, vol. 264, no. 2, pp. 665–694.
3. *Garey M. and Johnson D.* Computers and Intractability. N. Y., Freeman & Co, 1979. 340 p.
4. *Impagliazzo R. and Wigderson A.* P = BPP unless E has subexponential circuits: Derandomizing the XOR Lemma. Proc. 29th STOC, El Paso, ACM, 1997, pp. 220–229.
5. *Knuth D. E.* The Art of Computer Programming. Reading, Massachusetts, Addison-Wesley, 1997.
6. *Rybalov A.* O genericheskoy slozhnosti problemy obshcheznachimosti bulevykh formul [On generic complexity of the validity problem for Boolean formulas]. Prikladnaya Diskretnaya Matematika, 2016, no. 2(32), pp. 119–126. (in Russian)

# СВЕДЕНИЯ ОБ АВТОРАХ

**АЛЕКСЕЕВ Евгений Константинович** — кандидат физико-математических наук, начальник отдела криптографических исследований ООО «КРИПТО-ПРО», г. Москва. E-mail: **alekseev@cryptopro.ru**

**АЛЕХИНА Марина Анатольевна** — доктор физико-математических наук, профессор, заведующая кафедрой математики и физики Пензенского государственного технологического университета, г. Пенза. E-mail: **alekhina.marina19@yandex.ru, alekhina@penzgtu.ru**

**АХМЕТЗЯНОВА Лилия Руслановна** — заместитель начальника отдела криптографических исследований ООО «КРИПТО-ПРО», г. Москва.
E-mail: **lah@cryptopro.ru**

**БАБУЕВА Александра Алексеевна** — инженер-аналитик 1 категории отдела криптографических исследований ООО «КРИПТО-ПРО», г. Москва.
E-mail: **babueva@cryptopro.ru**

**КУЦЕНКО Александр Владимирович** — аспирант механико-математического факультета Новосибирского государственного университета, младший научный сотрудник Института математики им. С. Л. Соболева СО РАН, г. Новосибирск.
E-mail: **alexandrkutsenko@bk.ru**

**МИРОНКИН Владимир Олегович** — старший преподаватель кафедры компьютерной безопасности Московского института электроники и математики им. А. Н. Тихонова Национального исследовательского университета «Высшая школа экономики», г. Москва. E-mail: **mironkin.v@mail.ru**

**МОНАХОВА Эмилия Анатольевна** — кандидат технических наук, доцент, старший научный сотрудник Института вычислительной математики и математической геофизики СО РАН, г. Новосибирск. E-mail: **emilia@rav.sscc.ru**

**ОБЛАУХОВ Алексей Константинович** — аспирант, младший научный сотрудник института математики им. С. Л. Соболева, ассистент кафедры теоретической кибернетики Новосибирского государственного университета, исследователь Лаборатории криптографии JetBrains Research, г. Новосибирск. E-mail: **oblaukhov@gmail.com**

**ПЕРОВ Артём Андреевич** — ассистент кафедры информационной безопасности Московского политехнического университета, г. Москва.
E-mail: **perov_artem@inbox.ru**

**ПЕСТУНОВ Андрей Игоревич** — кандидат физико-математических наук, доцент, заведующий кафедрой информационных технологий Новосибирского государственного университета экономики и управления «НИНХ», г. Новосибирск.
E-mail: **pestunov@gmail.com**

**РЫБАЛОВ Александр Николаевич** — кандидат физико-математических наук, старший научный сотрудник лаборатории комбинаторных и вычислительных методов алгебры и логики Института математики им. С. Л. Соболева СО РАН, г. Омск.
E-mail: **alexander.rybalov@gmail.com**

**СМЫШЛЯЕВ Станислав Витальевич** — кандидат физико-математических наук, заместитель генерального директора ООО «КРИПТО-ПРО», г. Москва.
E-mail: **svs@cryptopro.ru**

**ТОКАРЕВА Наталья Николаевна** — кандидат физико-математических наук, доцент, старший научный сотрудник Института математики им. С. Л. Соболева СО РАН, г. Новосибирск. E-mail: **tokareva@math.nsc.ru**

**ЧЕРЕМУШКИН Александр Васильевич** — доктор физико-математических наук, профессор, член-корреспондент Академии криптографии РФ, научный консультант НИИ «Квант», г. Москва. E-mail: **avc238@mail.ru**