МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ

ПРИКЛАДНАЯ ДИСКРЕТНАЯ МАТЕМАТИКА

Научный журнал

2016 № 2(32)

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

ТОМСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

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

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

Периодичность выхода журнала: 4 номера в год.

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

Подписано к печати 15.06.2016.
Формат $60 \times 84\frac{1}{8}$. Усл. п. л. 13,4. Уч.-изд. л. 15. Тираж 300 экз. Заказ № 1918.

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

СОДЕРЖАНИЕ

ТЕОРЕТИЧЕСКИЕ ОСНОВЫ ПРИКЛАДНОЙ ДИСКРЕТНОЙ МАТЕМАТИКИ

МАТЕМАТИЧЕСКИЕ МЕТОДЫ КРИПТОГРАФИИ

МАТЕМАТИЧЕСКИЕ МЕТОДЫ СТЕГАНОГРАФИИ

ПРИКЛАДНАЯ ТЕОРИЯ ГРАФОВ

МАТЕМАТИЧЕСКИЕ ОСНОВЫ ИНФОРМАТИКИ И ПРОГРАММИРОВАНИЯ

CONTENTS

ТЕОРЕТИЧЕСКИЕ ОСНОВЫ ПРИКЛАДНОЙ ДИСКРЕТНОЙ МАТЕМАТИКИ

УДК 519.213

АССОЦИАТИВНЫЕ ФУНКЦИИ ФРАНКА В ПОСТРОЕНИИ СЕМЕЙСТВ ДИСКРЕТНЫХ ВЕРОЯТНОСТНЫХ РАСПРЕДЕЛЕНИЙ СЛУЧАЙНЫХ МНОЖЕСТВ СОБЫТИЙ

Н. А. Лукьянова, Д. В. Семенова

Институт математики и фундаментальной информатики Сибирского федерального университета, г. Красноярск, Россия

Работа является продолжением исследования проблемы рекуррентного построения класса дискретных вероятностных распределений случайного множества на конечном множестве из N событий. В качестве инструмента построения таких распределений предлагается использовать однопараметрическое семейство ассоциативных функций Франка. Исследуются его свойства и характеристики применительно к вероятностному распределению случайных множеств событий. Приводятся условия построения и существования полученных вероятностных распределений случайных множеств событий, а также их вид.

Ключевые слова: *случайное множество событий, дискретное вероятностное распределение, ассоциативная функция Франка.*

DOI 10.17223/20710410/32/1

ASSOCIATIVE FRANK FUNCTIONS IN CONSTRUCTING FAMILIES OF DISCRETE PROBABILITY DISTRIBUTIONS OF RANDOM SETS OF EVENTS

N. A. Lukyanova, D. V. Semenova

School of Mathematics and Computer Science, Siberian Federal University, Krasnoyarsk, Russia

E-mail: nata00sfu@gmail.com, dariasdv@gmail.com

Discrete probability distributions of random subsets on a finite set of events are considered. A one-parameter family of Frank associative functions is applied for generating them. The related properties and characteristics of functions in this family are described. The form and the creation and existence conditions of obtained distributions are also described.

Keywords: *random set of events, discrete probability distribution, associative function of Frank.*

Введение

В последние годы в многомерном анализе данных при моделировании случайных объектов нечисловой природы возрос интерес к изучению частично определённых моделей, в которых объектом статистического интереса является множество, а не точка.

Подобные объекты появились давно в статистике и эконометрике и описываются естественным образом случайными множествами [1, 2]. Основным направлением современной теории случайных множеств является разработка математического аппарата для описания характеристик случайного множества и процедур получения множества с желаемыми свойствами.

Центральным объектом нашего исследования является специфическое случайное множество, а именно — случайное конечное множество событий. Случайные множества событий позволяют выявить общие статистические закономерности распределения событий в различных системах объектов нечисловой природы.

В качестве математического аппарата для описания структуры зависимостей множества событий выступает вероятностное распределение случайного множества событий. Одной из главных проблем исследования случайных множеств событий является задача построения их вероятностных распределений, описывающих все способы взаимодействия элементов между собой в моделируемом множестве. Однако на пути решения этой задачи стоит известная преграда «проклятия размерности», которая заключается в экспоненциальном росте размерности вероятностного распределения случайного множества событий при возрастании числа событий, образующих это множество.

В [3 – 5] предложено организовать процесс построения вероятностных распределений случайных множеств событий на основе рекуррентного соотношения, полученного с помощью аппарата ассоциативных функций [6]. Основная идея построения — выразить вероятности пересечений множества событий функционально через вероятности самих событий, что приводит к уменьшению числа параметров, необходимых для построения вероятностных распределений случайных множеств событий. Данная работа является продолжением исследования проблемы рекуррентного построения класса дискретных вероятностных распределений случайного множества на конечном множестве из N событий. Рассматривается однопараметрическое семейство функций Франка, введённое в 1979 г. [7]. Исследуются его свойства и характеристики применительно к вероятностному распределению случайных множеств событий.

Организация работы следующая. В п. 1 приводятся основные сведения о ключевых объектах исследования — случайных множествах событий и вероятностных распределениях, их характеризующих, и инструменте исследования — аппарате ассоциативных функций; рассматривается рекуррентный метод построения сет-функций и исследуются свойства полученных сет-функций. В п. 2 исследуются ассоциативные функции Франка в рекуррентном построении семейств вероятностных распределений случайных множеств событий. Приводятся теоремы, устанавливающие условия построения, а также вид и условия существования полученных вероятностных распределений случайных множеств событий.

1. Основные понятия и обозначения

1.1. Случайное множество событий

Рассмотрим вероятностное пространство $(\Omega, \mathcal{F}, \mathbf{P})$. Пусть $\mathfrak{X} \subset \mathcal{F}$ — конечное множество событий, выбранных из алгебры $\mathcal{F}$ этого пространства, $|\mathfrak{X}| < \infty$.

Определение 1. Случайное множество событий K на конечном множестве событий $\mathfrak{X} \subset \mathcal{F}$ определяется как отображение $K : \Omega \to 2^{\mathfrak{X}}$, измеримое относительно пары алгебр $\left(\mathcal{F}, 2^{2^{\mathfrak{X}}}\right)$ в том смысле, что для всякого $X \in 2^{2^{\mathfrak{X}}}$ существует прообраз $K^{-1}(X) \in \mathcal{F}$, такой, что $\mathbf{P}(X) = \mathbf{P}(K^{-1}(X))$.

Замечание 1. Выражение $K(\omega) = \{x \in \mathfrak{X} : \omega \in x\}$ может быть истолковано как «случайное множество наступивших событий», поскольку элементарному исходу эксперимента $\omega \in \Omega$ ставится в соответствие некоторое подмножество событий $X \subseteq \mathfrak{X}$, которое содержит все те события, которые наступили в данном испытании.

Определение 2. Сет-функция $f(X)$, $X \in 2^{\mathfrak{X}}$, заданная на конечном множестве $\mathfrak{X}$, есть отображение $f : 2^{\mathfrak{X}} \to \mathbb{R}$.

Свойства сет-функций и условия их использования в качестве вероятностных распределений случайных множеств событий описаны в работах О. Ю. Воробьёва [8 – 10].

Случайное множество событий K, заданное на конечном множестве событий $\mathfrak{X}$, определяется своим вероятностным распределением. Если мощность рассматриваемого множества событий $|\mathfrak{X}| = N < \infty$, то имеется 2^N видов вероятностных зависимостей между событиями этого множества, т. е. ровно столько, сколько у этого множества подмножеств.

Определение 3. *Вероятностное распределение* случайного множества событий K, заданного на конечном множестве избранных событий $\mathfrak{X} \subset \mathcal{F}$, есть набор 2^N значений вероятностной меры $\mathbf{P}$ на событиях из $2^{\mathfrak{X}}$.

Замечание 2. Вероятностное распределение случайного множества событий K, заданного на конечном множестве избранных событий $\mathfrak{X} \subset \mathcal{F}$, является сет-функцией со значениями в $[0, 1]$.

Как известно, вероятностное распределение случайного множества событий можно задать шестью эквивалентными способами, которые определяются свойствами соответствующих сет-функций [8, 10]. В данной работе исследуются только два из них:

PI. Вероятностное распределение I рода случайного множества событий K на $\mathfrak{X}$ — это аддитивная сет-функция [8, 10], представляющая набор $\{p(X), X \subseteq \mathfrak{X}\}$ из 2^N вероятностей вида

$$p(X) = \mathbf{P}(\{K = X\}) = \mathbf{P}\left(\left(\bigcap_{x \in X} x\right) \bigcap \left(\bigcap_{x \in X^c} x^c\right)\right),$$

где $X^c = \mathfrak{X} \setminus X$, $x^c = \Omega \setminus x$, и удовлетворяющая следующим условиям:

$$0 \leqslant p(X) \leqslant 1, \quad X \subseteq \mathfrak{X}, \quad \sum_{X \subseteq \mathfrak{X}} p(X) = 1.$$

PII. Вероятностное распределение II рода случайного множества событий K на $\mathfrak{X}$ — это супераддитивная сет-функция [8, 10], представляющая набор из 2^N вероятностей вида $\{p_X, X \subseteq \mathfrak{X}\}$, где

$$p_X = \mathbf{P}(\{X \subseteq K\}) = \mathbf{P}\left(\left(\bigcap_{x \in X} x\right) \bigcap \left(\bigcap_{x \in X^c} \Omega\right)\right) = \mathbf{P}\left(\bigcap_{x \in X} x\right),$$

которые удовлетворяют системе неравенств Фреше:

$$0 \leqslant \max\left\{0, 1 - \sum_{x \in X} (1 - \mathbf{P}(x))\right\} \leqslant p_X \leqslant \min_{x \in X} \mathbf{P}(x) \leqslant 1.$$

Замечание 3. Для сокращения записи будем использовать следующие обозначения: $p_{\{x\}} = p_x$, $p(\{x\}) = p(x)$, $p_{\{x,y\}} = p_{xy}$, $p(\{x, y\}) = p(xy)$ и т. д.

Замечание 4. В теории случайных событий [3–5, 8, 10] обозначение $\varnothing$ используется для $\mathfrak{X}^c = \Omega \setminus \mathfrak{X}$. Событие $\{K = \varnothing\} = \bigcap\limits_{x\in\mathfrak{X}} x^c$ означает, что не наступило ни одно событие из $\mathfrak{X}$. По определению принимаем $\bigcap\limits_{x\in\varnothing} x = \Omega$ и $\bigcap\limits_{x\in\varnothing} x^c = \Omega$. Таким образом, в вероятностном распределении II рода всегда

$$p_\varnothing = \mathbf{P}(\{K \supseteq \varnothing\}) = \mathbf{P}\left(\left(\bigcap_{x\in\varnothing} x\right) \cap \left(\bigcap_{x\in\mathfrak{X}} \Omega\right)\right) = \mathbf{P}\left(\bigcap_{x\in\mathfrak{X}} \Omega\right) = \mathbf{P}(\Omega) = 1.$$

Вероятностные распределения I и II рода связаны взаимно-обратными формулами обращения Мёбиуса [8, 10] для всех $X \in 2^{\mathfrak{X}}$:

$$\begin{aligned} p_X &= \sum_{Y\in 2^{\mathfrak{X}}: X\subseteq Y} p(Y), \\ p(X) &= \sum_{Y\in 2^{\mathfrak{X}}: X\subseteq Y} (-1)^{|Y|-|X|} p_Y. \end{aligned} \tag{1}$$

Соотношению вероятностной нормировки удовлетворяет только вероятностное распределение I рода $\{p(X), X \subseteq \mathfrak{X}\}$, поскольку это вероятностный набор от полной группы несовместных событий $\left(\bigcap\limits_{x\in X} x\right) \cap \left(\bigcap\limits_{x\in X^c} x^c\right)$, образующих разбиение пространства элементарных исходов. Таким образом, если задано распределение I рода, то по формулам обращения Мёбиуса мы всегда получим распределение II рода $\left\{\mathbf{P}\left(\bigcap\limits_{x\in X} x\right),\ X \subseteq \mathfrak{X}\right\}$. Обратное не всегда верно, т. е. набор из 2^N чисел $\mathbf{P}(X)$, $X \subseteq \mathfrak{X}$, из $[0, 1]$, удовлетворяющих границам Фреше, не всегда определяет вероятностное распределение случайного множества событий.

Лемма 1. Если аддитивная сет-функция $f : 2^{\mathfrak{X}} \to [0,\ 1]$, заданная на значениях $\{K = X\} = \left(\bigcap\limits_{x\in X} x\right) \bigcap \left(\bigcap\limits_{x\in X^c} x^c\right)$, удовлетворяет условию $\sum\limits_{X\subseteq\mathfrak{X}} f(X) = 1$, то $f(X)$ определяет вероятностное распределение I рода случайного множества K на $\mathfrak{X}$.

Теорема 1. Если супераддитивная сет-функция $f : 2^{\mathfrak{X}} \to [0,\ 1]$, $X \subseteq \mathfrak{X}$, заданная на значениях $\{K \supseteq X\} = \left(\bigcap\limits_{x\in X} x\right)$, удовлетворяет следующим условиям:

— значение сет-функции на $\{K \supseteq \varnothing\}$ равно единице, т. е. $f(\varnothing) = 1$;
— значения сет-функции на событиях $x \in \mathfrak{X}$ совпадают с вероятностью этих событий, т. е. $f(x) = \mathbf{P}(x)$;
— значения сет-функции удовлетворяют системе неравенств Фреше

$$\max\left\{0, 1 - \sum_{x\in X}(1 - \mathbf{P}(x))\right\} \leqslant f(X) \leqslant \min_{x\in X} \mathbf{P}(x), \quad X \subseteq \mathfrak{X}, \quad |X| \geqslant 2;$$

— сет-функция $p(X)$, полученная по формулам обращения Мёбиуса

$$p(X) = \sum_{Y\in 2^{\mathfrak{X}}: X\subseteq Y} (-1)^{|Y|-|X|} f(Y),$$

определяет вероятностное распределение I рода,

то $f(X)$ определяет вероятностное распределение II рода случайного множества событий K на $\mathfrak{X}$.

1.2. Рекуррентное построение вероятностных распределений случайных множеств событий ассоциативными функциями

В современных теориях неопределённости широкое распространение получили классы ассоциативных функций [6]. Приведём определение ассоциативной функции, используемое в работе.

Определение 4. Ассоциативная функция $\mathrm{AF} : [0,1]^2 \to [0,1]$ определяется как двуместная функция, удовлетворяющая следующим свойствам:

A1. *Граничные условия:* $\mathrm{AF}(a,0) = \mathrm{AF}(0,a) = 0$, $\mathrm{AF}(a,1) = \mathrm{AF}(1,a) = a$, $a \in [0,1]$.
A2. *Монотонность:* $\forall a_1, a_2, b_1, b_2 \in [0,1] (a_1 \leqslant a_2 \& b_1 \leqslant b_2 \Rightarrow \mathrm{AF}(a_1,b_1) \leqslant \mathrm{AF}(a_2,b_2))$.
A3. *Коммутативность:* $\forall a, b \in [0,1]\ (\mathrm{AF}(a,b) = \mathrm{AF}(b,a))$.
A4. *Ассоциативность:* $\forall a, b, c \in [0,1]\ (\mathrm{AF}(\mathrm{AF}(a,b),c) = \mathrm{AF}(a,\mathrm{AF}(b,c)))$.
A5. *Условие липшиц-непрерывности:* $\forall a, b, c \in [0,1]\ (a \leqslant c \Rightarrow \mathrm{AF}(c,b) - \mathrm{AF}(a,b))$.

В [3–5] предложен рекуррентный подход к построению вероятностных распределений случайных множеств событий ассоциативными функциями. Основная идея этого подхода заключается в том, чтобы, исходя из известных вероятностей событий $\mathbf{P}(x) = p_x$, $x \in \mathfrak{X}$, формирование вероятностей пересечений событий $\mathbf{P}\left(\bigcap\limits_{x\in X} x\right)$ для $X \subseteq \mathfrak{X}$, $|X| > 1$, осуществлять последовательно согласно рекуррентной формуле

$$\mathbf{P}\left(\bigcap_{x\in X} x\right) = \mathrm{AF}\left(p_x, \mathbf{P}\left(\bigcap_{y\in X\setminus\{x\}} y\right)\right) = \ldots = \mathrm{AF}\left(p_x, x \in X\right). \tag{2}$$

Формула (2) позволяет функционально построить вероятность пересечения событий $X \subseteq \mathfrak{X}$, используя в качестве входных параметров N вероятностей событий и вид ассоциативной функции. В результате формируются $2^N - N - 1$ вероятностей пересечений событий, удовлетворяющих границам Фреше. Определим сет-функцию $\{f(X),\ X \subseteq \mathfrak{X}\}$ на конечном множестве событий $\mathfrak{X}$, значения которой определяются следующим образом:

$$\begin{array}{l} f(\varnothing) = 1; \\ f(x) = p_x = \mathbf{P}(x),\ x \in \mathfrak{X}; \\ f(X) = \mathrm{AF}\left(p_x, x \in X\right),\ X \subseteq \mathfrak{X},\ |X| > 1. \end{array} \tag{3}$$

Лемма 2. Пусть $\mathfrak{X} \subset \mathcal{F}$ — конечное множество событий, выбранных из алгебры $\mathcal{F}$ вероятностного пространства $(\Omega, \mathcal{F}, \mathbf{P})$, и известны вероятности событий $p_x = \mathbf{P}(x) > 0$, $x \in \mathfrak{X}$. Тогда для сет-функции $\{f(X), X \subseteq \mathfrak{X}\}$, значения которой определяются по формуле (3), справедливы $2^N - N - 1$ неравенств

$$0 \leqslant \mathrm{AF}\left(p_t, x \in X\right) \leqslant \mathrm{AF}\left(p_x, x \in X\right) \leqslant \mathrm{AF}(p_t, 1) = p_t \leqslant 1,\ \ X \subseteq \mathfrak{X},\ |X| > 2,$$

где $p_t = \min\limits_{x\in X} p_x$.

Лемма непосредственно следует из определения ассоциативной функции.

Следствие 1. Значения сет-функции, определяемой по формуле (3), на всех подмножествах $X \subseteq \mathfrak{X}$, $|X| > 2$, удовлетворяют неравенствам

$$\mathrm{AF}\left(p_x, x \in X\right) \leqslant \mathrm{AF}\left(p_x, x \in X \setminus \{t\}\right),\ t \in X.$$

Следствие вытекает из свойства A2 монотонности ассоциативной функции.

Заметим, что построенная сет-функция является вероятностным распределением случайного множества событий при выполнении условий теоремы 1. Следовательно, для каждого семейства ассоциативных функций [6] необходимо определять условия, при которых сет-функция (3) является вероятностным распределением случайного множества событий. Для этого необходимо по формулам обращения Мёбиуса (1) перейти к вероятностному распределению I рода и определить условия, при которых выполняются свойства неотрицательности и нормировки.

2. Ассоциативные функции Франка

Рассмотрим однопараметрическое семейство функций Франка [7]

$$\mathrm{Frank}(a,b;\alpha)=\mathrm{AF}_\alpha(a,b)=\ln\left(1+\frac{(e^{-\alpha\cdot a}-1)(e^{-\alpha\cdot b}-1)}{(e^{-\alpha}-1)}\right)^{-1/\alpha},\quad \alpha\neq 0, \tag{4}$$

где α представляет собой параметр, определяющий зависимость между переменными a и b, в качестве которых будем рассматривать вероятности событий $\mathbf{P}(x)=p_x=a$ и $\mathbf{P}(y)=p_y=b$.

Далее применим рекуррентный метод [3 – 5] с функцией Франка (4) для построения вероятностного распределения случайного множества событий. Сначала найдём общий вид сет-функции, а затем определим условия, при которых построенная сет-функция будет вероятностным распределением II рода некоторого случайного множества событий, заданного на конечном множестве $\mathfrak{X}$.

Лемма 3. Для значений сет-функции $\{f(X), X\subseteq\mathfrak{X}\}$, построенной рекуррентным методом (3) с ассоциативной функцией (4), справедлива формула

$$f(X)=-\frac{1}{\alpha}\ln\left(1+\frac{\prod\limits_{x\in X}(e^{-\alpha p_x}-1)}{(e^{-\alpha}-1)^{|X|-1}}\right),\quad X\subseteq\mathfrak{X},\ \alpha\neq 0. \tag{5}$$

Доказательство. Будем доказывать формулу (5) для подмножеств X из $\mathfrak{X}$, рассматривая их по возрастанию мощности ($|X|=0,1,\dots N$) и находя значения сет-функции на $X\subseteq\mathfrak{X}$, используя рекуррентный метод (3) с ассоциативной функцией Франка (4).

Пусть $|X|=0$. Согласно рекуррентному методу, полагаем значение сет-функции на пустом множестве событий равным единице: $f(\varnothing)=1$.

Пусть $|X|=1$. Тогда для любого монополета $X=\{x\}$, состоящего из одного события $x\in\mathfrak{X}$, полагаем значения сет-функции на событиях равными вероятностям этих событий, т. е. $f(X)=\mathbf{P}(x)=p_x$.

Пусть $|X|=2$. Рассмотрим произвольный дуплет событий $X=\{x,y\}$, $x,y\in\mathfrak{X}$. Согласно (2) и (4), значение сет-функции равно

$$f(\{x,y\})=\mathbf{P}(x\cap y)=\mathrm{Frank}(p_x,p_y;\alpha)=-\frac{1}{\alpha}\ln\left(1+\frac{(e^{-\alpha p_x}-1)(e^{-\alpha p_y}-1)}{(e^{-\alpha}-1)}\right).$$

Пусть $|X|=3$. Рассмотрим произвольный триплет событий $X=\{x,y,z\}$, $x,y,z\in$ $\in\mathfrak{X}$. Найдём значение сет-функции на этом триплете, используя (2), (4) и значение

сет-функции $f(\{x,y\})$, полученное на предыдущем шаге:

$$f(\{x,y,z\}) = \mathbf{P}(x \cap y \cap z) = \mathrm{Frank}(\mathrm{Frank}(p_x,p_y;\alpha),p_z;\alpha) =$$
$$= \mathrm{Frank}(f(\{x,y\}),p_z;\alpha) = -\frac{1}{\alpha}\ln\left(1+\frac{(e^{-\alpha\cdot f(\{x,y\})}-1)(e^{-\alpha p_z}-1)}{(e^{-\alpha}-1)}\right). \tag{6}$$

Рассмотрим $e^{-\alpha\cdot f(\{x,y\})}$:

$$e^{-\alpha\cdot f(\{x,y\})} = e^{-\alpha\left[-\frac{1}{\alpha}\ln\left(1+\frac{(e^{-\alpha p_x}-1)(e^{-\alpha p_y}-1)}{(e^{-\alpha}-1)}\right)\right]} = 1+\frac{(e^{-\alpha p_x}-1)(e^{-\alpha p_y}-1)}{(e^{-\alpha}-1)}. \tag{7}$$

Подставим (7) в (6) и получим

$$f(\{x,y,z\}) = -\frac{1}{\alpha}\ln\left(1+\frac{\left(\left(1+\frac{(e^{-\alpha p_x}-1)(e^{-\alpha p_y}-1)}{(e^{-\alpha}-1)}\right)-1\right)(e^{-\alpha p_z}-1)}{(e^{-\alpha}-1)}\right) =$$
$$= -\frac{1}{\alpha}\ln\left(1+\frac{(e^{-\alpha p_x}-1)(e^{-\alpha p_y}-1)(e^{-\alpha p_z}-1)}{(e^{-\alpha}-1)^2}\right).$$

Пусть $|X|=4$. Рассмотрим произвольный четырёхплет событий $X=\{x,y,z,v\}$, $x,y,z,v\in\mathfrak{X}$. Найдём значение сет-функции на этом четырёхплете, используя (2), (4) и значение сет-функции $f(\{x,y,z\})$, полученное на предыдущем шаге для триплета событий:

$$f(\{x,y,z,v\}) = \mathbf{P}(x\cap y\cap z\cap v) =$$
$$= \mathrm{Frank}(\mathrm{Frank}(\mathrm{Frank}(p_x,p_y;\alpha),p_z;\alpha),p_v;\alpha) = \mathrm{Frank}(f(\{x,y,z\}),p_v;\alpha) =$$
$$= -\frac{1}{\alpha}\ln\left(1+\frac{(e^{-\alpha f(\{x,y,z\})}-1)(e^{-\alpha p_v}-1)}{(e^{-\alpha}-1)}\right).$$

Аналогично (7) можно доказать, что

$$e^{-\alpha f(\{x,y,z\})} = 1+\frac{(e^{-\alpha p_x}-1)(e^{-\alpha p_y}-1)(e^{-\alpha p_z}-1)}{(e^{-\alpha}-1)^2}.$$

Следовательно,

$$f(\{x,y,z,v\}) = -\frac{1}{\alpha}\ln\left(1+\frac{\left[1+\frac{(e^{-\alpha p_x}-1)(e^{-\alpha p_y}-1)(e^{-\alpha p_z}-1)}{(e^{-\alpha}-1)^2}-1\right](e^{-\alpha p_v}-1)}{(e^{-\alpha}-1)}\right) =$$
$$= -\frac{1}{\alpha}\ln\left(1+\frac{(e^{-\alpha p_x}-1)(e^{-\alpha p_y}-1)(e^{-\alpha p_z}-1)(e^{-\alpha p_v}-1)}{(e^{-\alpha}-1)^3}\right).$$

Пусть $|X|=n$, $n\leqslant N$. Для произвольного n-плета событий X значения сет-функции будем находить аналогичным образом, используя рекуррентную формулу (2) и значения сет-функции, найденные на предыдущих $n-1$ шагах:

$$f(X) = \mathbf{P}\left(\bigcap_{x\in X} x\right) = \mathrm{Frank}\left(f(X\setminus\{t\}),p_t;\alpha\right) = -\frac{1}{\alpha}\ln\left(1+\frac{\prod\limits_{x\in X}(e^{-\alpha p_x}-1)}{(e^{-\alpha}-1)^{n-1}}\right).$$

Лемма доказана. ■

Рассмотрим условия, при которых построенная сет-функция $\{f(X), X\subseteq\mathfrak{X}\}$ является вероятностным распределением случайного множества.

Теорема 2. Пусть $\mathfrak{X} \subset \mathcal{F}$, $N = |\mathfrak{X}|$, — конечное множество событий, выбранных из алгебры $\mathcal{F}$ вероятностного пространства $(\Omega, \mathcal{F}, \mathbf{P})$, и известны вероятности событий $p_x = \mathbf{P}(x) > 0$, $x \in \mathfrak{X}$. Тогда сет-функция $\{f(X), X \subseteq \mathfrak{X}\}$, значения которой определяются рекуррентным методом с ассоциативной функцией Франка (4), является вероятностным распределением II рода случайного множества событий при выполнении следующей системы из $2^N - N - 1$ неравенств:

$$1 \leqslant \prod_{Y \in 2^{\mathfrak{X}}: X \subseteq Y} \left(1 + \frac{\prod\limits_{x \in Y} (e^{-\alpha p_x} - 1)}{(e^{-\alpha} - 1)^{|Y|-1}}\right)^{\frac{(-1)^{|Y|-|X|+1}}{\alpha}} \leqslant e, \quad X \subseteq \mathfrak{X}, \quad \alpha \neq 0.$$

Доказательство. Согласно лемме 3, значения сет-функции $\{f(X), X \subseteq \mathfrak{X}\}$, построенной рекуррентным методом с ассоциативной функцией (4), определяются по формуле (5). Построим новую сет-функцию $\{p(X), X \subseteq \mathfrak{X}\}$, преобразовав сет-функцию $\{f(X), X \subseteq \mathfrak{X}\}$ по формулам обращения Мёбиуса (1). Напомним, что сет-функция $f(X)$ задана на множествах событий $\{K \supseteq X\} = \left(\bigcap\limits_{x \in X} x\right)$, а сет-функция $p(X)$ — на множествах событий $\{K = X\} = \left(\bigcap\limits_{x \in X} x\right) \cap \left(\bigcap\limits_{x \in X^c} x^c\right)$ [8, 10].

Для любого $X \subseteq \mathfrak{X}$ по формуле (1) получаем

$$p(X) = \sum_{Y \in 2^{\mathfrak{X}}: X \subseteq Y} (-1)^{|Y|-|X|} \cdot f(Y) = \sum_{Y \in 2^{\mathfrak{X}}: X \subseteq Y} (-1)^{|Y|-|X|} \cdot \ln \left(1 + \frac{\prod\limits_{x \in Y} (e^{-\alpha p_x} - 1)}{(e^{-\alpha} - 1)^{|Y|-1}}\right)^{\frac{1}{\alpha}} =$$

$$= \ln \prod_{Y \in 2^{\mathfrak{X}}: X \subseteq Y} \left(1 + \frac{\prod\limits_{x \in Y} (e^{-\alpha p_x} - 1)}{(e^{-\alpha} - 1)^{|Y|-1}}\right)^{\frac{(-1)^{|Y|-|X|+1}}{\alpha}}.$$

Таким образом, получили новую сет-функцию $\{p(X), X \subseteq \mathfrak{X}\}$ на множестве событий $\mathfrak{X}$ со значениями, которые определяются по формуле

$$p(X) = \ln \prod_{Y \in 2^{\mathfrak{X}}: X \subseteq Y} \left(1 + \frac{\prod\limits_{x \in Y} (e^{-\alpha p_x} - 1)}{(e^{-\alpha} - 1)^{|Y|-1}}\right)^{\frac{(-1)^{|Y|-|X|+1}}{\alpha}}, \quad X \subseteq X. \tag{8}$$

Полученная сет-функция является распределением I рода по лемме 1, если для неё выполняются условия неотрицательности $0 \leqslant p(X) \leqslant 1$, $X \subseteq \mathfrak{X}$ и $\sum\limits_{X \subseteq \mathfrak{X}} p(X) = 1$.

Выполнение последнего условия гарантирует формула обращения Мёбиуса (1) [8].

Проверим условия неотрицательности значений полученной сет-функции. Будем рассматривать значения по убыванию мощности, т. е. $|X| = N, N-1, \ldots, 0$.

Пусть $|X| = N$, т. е. $X = \mathfrak{X}$. Необходимо доказать, что $0 \leqslant p(\mathfrak{X}) \leqslant 1$. Из (8) имеем

$$p(\mathfrak{X}) = \ln \prod_{Y \in 2^{\mathfrak{X}}: \mathfrak{X} \subseteq Y} \left(1 + \frac{\prod\limits_{x \in Y} (e^{-\alpha p_x} - 1)}{(e^{-\alpha} - 1)^{|Y|-1}}\right)^{\frac{(-1)^{|Y|-N+1}}{\alpha}} = \ln \left(1 + \frac{\prod\limits_{x \in \mathfrak{X}} (e^{-\alpha p_x} - 1)}{(e^{-\alpha} - 1)^{N-1}}\right)^{-\frac{1}{\alpha}} = f(\mathfrak{X}).$$

Из леммы 2 следует, что $0 \leqslant p(\mathfrak{X}) \leqslant 1$.

Пусть $|X| = N-1$, т. е. $X = \mathfrak{X} \setminus \{t\}$. Из (8) имеем

$$p(\mathfrak{X}\setminus\{t\}) = \ln \prod_{Y\in 2^{\mathfrak{X}}:\mathfrak{X}\setminus\{t\}\subseteq Y} \left(1+\frac{\prod\limits_{x\in Y}(e^{-\alpha p_x}-1)}{(e^{-\alpha}-1)^{|Y|-1}}\right)^{\frac{(-1)^{|Y|-(N-1)+1}}{\alpha}} =$$

$$= \ln\left(1+\frac{\prod\limits_{x\in\mathfrak{X}\setminus\{t\}}(e^{-\alpha p_x}-1)}{(e^{-\alpha}-1)^{N-2}}\right)^{\frac{-1}{\alpha}} \cdot \left(1+\frac{\prod\limits_{x\in\mathfrak{X}}(e^{-\alpha p_x}-1)}{(e^{-\alpha}-1)^{N-1}}\right)^{\frac{1}{\alpha}} =$$

$$= \ln\left(1+\frac{\prod\limits_{x\in\mathfrak{X}\setminus\{t\}}(e^{-\alpha p_x}-1)}{(e^{-\alpha}-1)^{N-2}}\right)^{\frac{-1}{\alpha}} - \ln\left(1+\frac{\prod\limits_{x\in\mathfrak{X}}(e^{-\alpha p_x}-1)}{(e^{-\alpha}-1)^{N-1}}\right)^{\frac{-1}{\alpha}} = f(\mathfrak{X}\setminus\{t\}) - f(\mathfrak{X}).$$

Из леммы 2 и следствия 1 очевидно выполнение неравенства $0 \leqslant f(\mathfrak{X}\setminus\{t\}) - f(\mathfrak{X}) \leqslant 1$, отсюда $0 \leqslant p(\mathfrak{X}\setminus\{t\}) \leqslant 1$.

Получим оценку для оставшихся $2^N - N - 1$ множеств $X \subseteq \mathfrak{X}$ для (8):

$$0 \leqslant \ln \prod_{Y\in 2^{\mathfrak{X}}:X\subseteq Y} \left(1+\frac{\prod\limits_{x\in Y}(e^{-\alpha p_x}-1)}{(e^{-\alpha}-1)^{|Y|-1}}\right)^{\frac{(-1)^{|Y|-|X|+1}}{\alpha}} \leqslant 1. \tag{9}$$

Используя свойства логарифма, переходим к следующей оценке:

$$1 \leqslant \prod_{Y\in 2^{\mathfrak{X}}:X\subseteq Y} \left(1+\frac{\prod\limits_{x\in Y}(e^{-\alpha p_x}-1)}{(e^{-\alpha}-1)^{|Y|-1}}\right)^{\frac{(-1)^{|Y|-|X|+1}}{\alpha}} \leqslant e, \quad X \subseteq \mathfrak{X},\ \alpha \neq 0. \tag{10}$$

Следовательно, сет-функция, значения которой определяются рекуррентным методом с ассоциативной функцией Франка, является вероятностным распределением II рода случайного множества событий при выполнении системы из $2^N - N - 1$ неравенств (10). ■

Теорема 2 даёт инструмент для построения случайных множеств событий с заранее заданной структурой зависимостей. Из неё следует, что выполнение системы неравенств (10) зависит от выбора значения параметра α. Рассмотрим предельные случаи при $\alpha \to 0^{\pm}$ и $\alpha \to \pm\infty$.

Теорема 3. Пусть $\mathfrak{X} \subset \mathcal{F}$ — конечное множество событий, выбранных из алгебры $\mathcal{F}$ вероятностного пространства $(\Omega, \mathcal{F}, \mathbf{P})$; случайное множество событий K на конечном множестве событий $\mathfrak{X} \subset \mathcal{F}$ определяется вероятностным распределением II рода $\{p_X, X \subseteq \mathfrak{X}\}$, построенным ассоциативной функцией Франка

$$p_X = -\frac{1}{\alpha}\ln\left(1+\frac{\prod\limits_{x\in X}(e^{-\alpha p_x}-1)}{(e^{-\alpha}-1)^{|X|-1}}\right), \quad x \in X, \quad X \subseteq \mathfrak{X},\ \alpha \neq 0.$$

Тогда функция p_X

1) при $\alpha \to 0^{\pm}$ стремится к вероятностному распределению независимо-точечного случайного множества событий, определяемому ассоциативной функцией

$$\mathrm{AF}\,(p_x, x \in X) = \prod_{x \in X} p_x;$$

2) при $\alpha \to +\infty$ стремится к вероятностному распределению случайного множества вложенных событий, определяемому ассоциативной функцией

$$\mathrm{AF}\,(p_x, x \in X) = \min_{x \in X} p_x;$$

3) при $\alpha \to -\infty$ стремится к вероятностному распределению случайного множества событий, определяемому ассоциативной функцией

$$\mathrm{AF}\,(p_x, x \in \mathfrak{X}) = \max\left\{\sum_{x \in X} p_x - |X| + 1, 0\right\}, \quad |X| > 1.$$

Доказательство.

1) Рассмотрим вероятностное распределение случайного множества событий, построенное ассоциативной функцией Франка (4) при $\alpha \to 0^{\pm}$. Для всех $X \subseteq \mathfrak{X}$

$$\lim_{\alpha \to 0^{\pm}} p_X = \lim_{\alpha \to 0^{\pm}} \ln \left(1 + \frac{\prod\limits_{x \in X} (e^{-\alpha p_x} - 1)}{(e^{-\alpha} - 1)^{|X|-1}}\right)^{-\frac{1}{\alpha}} =$$

$$= \ln \lim_{\alpha \to 0^{\pm}} \left(1 + \frac{\prod\limits_{x \in X} (e^{-\alpha p_x} - 1)}{(e^{-\alpha} - 1)^{|X|-1}}\right)^{-\frac{1}{\alpha}} = \ln \lim_{\alpha \to 0^{\pm}} \left(1 + \frac{\prod\limits_{x \in X} (-\alpha p_x)}{(-\alpha)^{|X|-1}}\right)^{-\frac{1}{\alpha}} =$$

(на этом этапе используем эквивалентность бесконечно малых величин, т. е. для $\alpha \to 0^{\pm}$ имеет место $(e^{-\alpha p_x} - 1) \sim (-\alpha p_x)$, $(e^{-\alpha} - 1) \sim (-\alpha)$)

$$= \ln \lim_{\alpha \to 0^{\pm}} \left(1 + \frac{(-\alpha)^{|X|} \prod\limits_{x \in X} p_x}{(-\alpha)^{|X|-1}}\right)^{-\frac{1}{\alpha}} = \ln \lim_{\alpha \to 0^{\pm}} \left(1 + (-\alpha) \prod_{x \in X} p_x)\right)^{-\frac{1}{\alpha}} =$$

$$= (\text{применяем второй замечательный предел}) = \ln e^{\lim\limits_{\alpha \to 0^{\pm}} \prod\limits_{x \in X} p_x} = \ln e^{\prod\limits_{x \in X} p_x} = \prod_{x \in X} p_x.$$

Следовательно, при $\alpha \to 0^{\pm}$ вероятностное распределение, построенное ассоциативной функцией Франка (4), стремится к вероятностному распределению независимо-точечного случайного множества событий [3 – 5], определяемому ассоциативной функцией $\mathrm{AF}(a, b) = a \cdot b$.

2) Рассмотрим вероятностное распределение случайного множества событий, построенное ассоциативной функцией Франка (4), когда $\alpha \to +\infty$. Сначала выделим $p_t = \min\limits_{x \in X} p_X$:

$$-\frac{1}{\alpha} \ln \left(1 + \frac{\prod\limits_{x \in X} (e^{-\alpha p_x} - 1)}{(e^{-\alpha} - 1)^{|X|-1}}\right) = -\frac{1}{\alpha} \ln \left(1 + \frac{(e^{-\alpha p_t} - 1) \cdot \prod\limits_{x \in X \setminus \{t\}} (e^{-\alpha p_x} - 1)}{(e^{-\alpha} - 1)^{|X|-1}}\right) \leqslant$$

(теперь используем свойство монотонности A2, так как все $p_x \leqslant p_\varnothing = 1$, получаем)

$$\leqslant -\frac{1}{\alpha}\ln\left(1+\frac{(e^{-\alpha p_t}-1)\cdot\prod\limits_{x\in X\setminus\{t\}}(e^{-\alpha p_\varnothing}-1)}{(e^{-\alpha}-1)^{|X|-1}}\right)=$$

$$=-\frac{1}{\alpha}\ln\left(1+\frac{(e^{-\alpha p_t}-1)(e^{-\alpha}-1)^{|X|-1}}{(e^{-\alpha}-1)^{|X|-1}}\right)=p_t=\min_{x\in X}p_X.$$

Тогда

$$\lim_{\alpha\to+\infty}-\frac{1}{\alpha}\ln\left(1+\frac{\prod\limits_{x\in X}(e^{-\alpha p_x}-1)}{(e^{-\alpha}-1)^{|X|-1}}\right)\leqslant\min_{x\in X}p_X.$$

Итак, для $\alpha\to+\infty$ вероятностное распределение, построенное ассоциативной функцией Франка (4), стремится к вероятностному распределению по ассоциативной функции $\mathrm{AF}(a,b)=\min\{a,b\}$, которая определяет случайное множество вложенных событий с вероятностным распределением II рода [3, 4].

3) Рассмотрим вероятностное распределение случайного множества событий, построенное ассоциативной функцией Франка (4), когда $\alpha\to-\infty$.

Для $\mathbf{P}(x\cap y)=\mathrm{Frank}(p_x,p_y;\alpha)$, где $X=\{x,y\}$, $X\subseteq\mathfrak{X}$, получаем

$$\lim_{\alpha\to-\infty}\mathrm{Frank}(p_x,p_y;\alpha)=\lim_{\alpha\to-\infty}-\frac{1}{\alpha}\ln\left(1+\frac{(e^{-\alpha p_x}-1)(e^{-\alpha p_y}-1)}{e^{-\alpha}-1}\right)=$$

$$=\lim_{\alpha\to-\infty}-\frac{1}{\alpha}\ln\left(\frac{e^{-\alpha}+e^{-\alpha(p_x+p_y)}-e^{-\alpha p_x}-e^{-\alpha p_y}}{e^{-\alpha}-1}\right)=\lim_{\alpha\to-\infty}\frac{\ln\left(1+e^{-\alpha(p_x+p_y-1)}\right)}{-\alpha}=$$

$$=(\text{по правилу Лопиталя})=\lim_{\alpha\to-\infty}\frac{\dfrac{e^{-\alpha(p_x+p_y-1)}}{1+e^{-\alpha(p_x+p_y-1)}}(-(p_x+p_y-1))}{-1}=$$

$$=(p_x+p_y-1)\lim_{\alpha\to-\infty}\frac{e^{-\alpha(p_x+p_y-1)}}{1+e^{-\alpha(p_x+p_y-1)}}=\begin{cases}p_x+p_y-1, & \text{если}\quad p_x+p_y>1,\\ 0 & \text{иначе}\ .\end{cases}$$

Для $\mathbf{P}(x\cap y\cap z)=\mathrm{Frank}(\mathrm{Frank}(p_x,p_y,\alpha),p_z;\alpha)$, $X=\{x,y,z\}$, $X\subseteq\mathfrak{X}$, имеем

$$\lim_{\alpha\to-\infty}-\frac{1}{\alpha}\ln\left(1+\frac{(e^{-\alpha p_x}-1)(e^{-\alpha p_y}-1)(e^{-\alpha p_z}-1)}{(e^{-\alpha}-1)^2}\right)=$$

$$=\lim_{\alpha\to-\infty}-\frac{1}{\alpha}\ln\left(\frac{(e^{-\alpha}-1)^2+(e^{-\alpha p_x}-1)(e^{-\alpha p_y}-1)(e^{-\alpha p_z}-1)}{(e^{-\alpha}-1)^2}\right)=$$

$$=\lim_{\alpha\to-\infty}\frac{\ln\left(1+e^{-\alpha(p_x+p_y+p_z-2)}\right)}{-\alpha}=$$

$$=\left(\sum_{t\in X}p_t-2\right)\lim_{\alpha\to-\infty}\frac{e^{-\alpha(p_x+p_y+p_z-2)}}{1+e^{-\alpha(p_x+p_y+p_z-2)}}=\begin{cases}\sum\limits_{t\in X}p_t-2, & \text{если}\quad \sum\limits_{t\in X}p_t>2,\\ 0 & \text{иначе.}\end{cases}$$

...

Для n-плетов, $n < N$, $|X| = n$, получим

$$\lim_{\alpha\to-\infty} -\frac{1}{\alpha}\ln\left(1+\frac{\prod\limits_{t\in X}(e^{-\alpha p_t}-1)}{(e^{-\alpha}-1)^{n-1}}\right) = \lim_{\alpha\to-\infty} -\frac{1}{\alpha}\ln\left(\frac{(e^{-\alpha}-1)^{n-1}+\prod\limits_{t\in X}(e^{-\alpha p_t}-1)}{(e^{-\alpha}-1)^{n-1}}\right) =$$

$$= \lim_{\alpha\to-\infty} \frac{\ln\left(1+e^{-\alpha\left(\sum\limits_{t\in X} p_t-n+1\right)}\right)}{-\alpha} = \lim_{\alpha\to-\infty} \frac{\left(\ln\left(1+e^{-\alpha\left(\sum\limits_{t\in X} p_t-n+1\right)}\right)\right)'_\alpha}{-\alpha'_\alpha} =$$

$$= \left(\sum_{t\in X} p_t - n + 1\right) \lim_{\alpha\to-\infty} \frac{e^{-\alpha\left(\sum\limits_{t\in X} p_t-n+1\right)}}{1+e^{-\alpha\left(\sum\limits_{t\in X} p_t-n+1\right)}} = \begin{cases} \sum\limits_{t\in X} p_t - |X| + 1, & \text{если } \sum\limits_{t\in X} p_t > |X|-1, \\ 0 & \text{иначе.} \end{cases}$$

Итак, для $\alpha \to -\infty$ вероятностное распределение, построенное ассоциативной функцией Франка (4), стремится к вероятностному распределению по ассоциативной функции $\mathrm{AF}(a,b) = \max\{a+b-1, 0\}$. В [3, 4] показано, что вероятностное распределение случайного множества событий с использованием этой функции возможно только при выполнении определённых ограничений на входные вероятности событий. При этом возникают только три вида результирующих случайных множеств событий с соответствующими вероятностными распределениями:

1) случайное множество непересекающихся событий, если выполнено условие

$$\sum_{x\in\mathfrak{X}} p_x \leqslant 1, \quad p_X = 0, \quad |X| > 1;$$

2) случайное множество событий, принимающее значения с ненулевой вероятностью лишь на подмножествах мощности $|\mathfrak{X}|-1$ и $|\mathfrak{X}|$, если $|\mathfrak{X}|-1 < \sum\limits_{x\in\mathfrak{X}} p_x \leqslant |\mathfrak{X}|$;

3) случайное множество событий, принимающее значения с ненулевой вероятностью лишь на подмножествах мощности $|\mathfrak{X}|-1$, если $\sum\limits_{x\in\mathfrak{X}} p_x = |\mathfrak{X}|-1$.

Теорема доказана. ■

Для множества $\mathfrak{X}$, состоящего из N событий, существует набор из $2^N - N - 1$ арных ковариаций [10], которые определяются вероятностным распределением II рода случайного множества событий на $\mathfrak{X}$. $|X|$-арная ковариация произвольного множества событий $X \subset \mathcal{F}$, $|X| > 1$, с учётом (2) примет следующий вид:

$$\mathrm{Kov}_X = \mathbf{P}\left(\bigcap_{x\in X} x\right) - \prod_{x\in X} p_x = \mathrm{AF}\left(p_x, \mathbf{P}\left(\bigcap_{y\in X\setminus\{x\}} y\right)\right) - \prod_{x\in X} p_x = \mathrm{AF}\left(p_x, x\in X\right) - \prod_{x\in X} p_x.$$

Теорема 4. Для вероятностного распределения случайного множества событий, построенного ассоциативной функцией Франка, все арные ковариации Kov_X, $X \subseteq \mathfrak{X}$, $|X| > 1$, имеют вид

$$\mathrm{Kov}_X = -\frac{1}{\alpha}\ln\left(\left(1+\frac{\prod\limits_{x\in X}(e^{-\alpha p_x}-1)}{(e^{-\alpha}-1)^{|X|-1}}\right) e^{\alpha\cdot\prod\limits_{x\in X} p_x}\right), \quad X \subseteq \mathfrak{X}.$$

Знак ковариации определяется знаком параметра $\alpha \in (-\infty, +\infty) \setminus \{0\}$.

Доказательство. Обозначим $E_X = 1 + \dfrac{\prod\limits_{x\in X}(e^{-\alpha p_x}-1)}{(e^{-\alpha}-1)^{|X|-1}}$. Тогда

$$\mathrm{Kov}_X = p_X - \prod_{x\in X} p_x = -\tfrac{1}{\alpha}\ln(E_X) - \prod_{x\in X} p_x = -\frac{1}{\alpha}\ln(E_X) - \ln e^{\prod\limits_{x\in X} p_x} =$$

$$= -\frac{1}{\alpha}\ln(E_X) - \ln\left(\left(e^{\prod\limits_{x\in X} p_x}\right)^{\alpha}\right)^{\frac{1}{\alpha}} = -\frac{1}{\alpha}\ln(E_X) - \frac{1}{\alpha}\ln e^{\alpha\cdot\prod\limits_{x\in X} p_x} =$$

$$= -\frac{1}{\alpha}\ln\left(E_X\cdot e^{\alpha\cdot\prod\limits_{x\in X} p_x}\right) = -\frac{1}{\alpha}\ln\left(\left(1+\frac{\prod\limits_{x\in X}(e^{-\alpha p_x}-1)}{(e^{-\alpha}-1)^{|X|-1}}\right)e^{\alpha\cdot\prod\limits_{x\in X} p_x}\right).$$

Теорема доказана. ■

Заключение

Одной из важных, сложных и значимых задач для науки в целом и для отдельных сфер вероятностных приложений, таких, как экономика и статистика, является разработка методов определения, изучения и статистической оценки структуры зависимостей сложных распределений множеств событий большой размерности. Распределение случайного множества событий — это удобный математический аппарат для описания всех способов взаимодействия элементов между собой. Следует отметить, что вероятностное распределение случайного множества событий определяется набором 2^N параметров. Однако на практике из статистики нам доступны, как правило, лишь вероятности самих событий (N параметров). В работе рассматривается новый подход моделирования вероятностных распределений случайного множества событий, который позволяет уменьшить размерность задачи с 2^N до N параметров. Исследовано однопараметрическое семейство функций Франка в построении вероятностных распределений случайных множеств событий. Доказаны следующие теоремы:

— об условиях, при которых сет-функция, построенная ассоциативной функцией Франка, является вероятностным распределением II рода; о виде вероятностного распределения случайного множества событий I рода;
— о виде арных ковариаций.

Семейство всех подмножеств любого достаточно большого множества событий огромно, что приводит к невозможности определения нетривиального вероятностного распределения. В связи с этим необходим инструмент, который позволял бы аналитику «сортировать» вероятностные распределения случайного множества событий, учитывая их конкретные свойства. Использование однопараметрического семейства функций Франка позволяет получать набор распределений, описывающих случайные множества событий, структура зависимостей которых регулируется параметром α. Предложенный метод не претендует на универсальность, однако он позволяет получать в виде вероятностных распределений случайного множества событий и их характеристик входные данные для ряда моделей статистических систем, которые обладают сложной структурой зависимостей и взаимосвязей [10]. С помощью предложенного подхода можно решать задачу по восстановлению пропущенных множественных данных и прогнозирования числовых величин на основе нечисловой информации.

Случайные множества с распределениями, построенными ассоциативной функцией Франка, использовались в событийном анализе медицинских данных. На основе холтеровского мониторирования по данным за 2014–2015 г. Федерального государственного бюджетного учреждения «Федеральный Сибирский научно-клинический центр

Федерального медико-биологического агентства России» г. Красноярска было проанализировано множество событийных факторов «Патологии» = {наджелудочковые экстрасистолы; наджелудочковая тахикардия; желудочковые экстрасистолы; желудочковая тахикардия; AV-блокады; СА-блокады; внутрижелудочковые блокады; ишемия миокарда} и множество целевых событий «Целевые рекомендации» = {мероприятия по изменению образа жизни; медикаментозная терапия; кардиостимулятор; др. виды оперативного лечения}, характеризующее вмешательство врача для обеспечения повышения качества охраны здоровья пациента. По пациентам, для которых проводилось холтеровское исследование, сформировали группы по признакам: возраст, пол, район проживания. Данные группы определили множество «Пациенты». Главная цель обработки статистических данных состояла в обнаружении скрытых в них закономерностей. Эти закономерности или знания позволяют выявить и понять сущность зависимостей и, опираясь на имеющиеся данные, принимать управленческие решения. Для определения зависимости между двумя случайными множествами событий в [11] было предложено использовать сет-регрессию, которая устанавливает вид средней функциональной зависимости между этими двумя случайными множествами событий. Решение задачи сет-регрессии требует знания совместного распределения случайных множеств событий K и L, значения которых для данного примера содержатся в конечных множествах «Патологии» и «Целевые рекомендации». Совместное распределение формировалось с помощью ассоциативной функции Франка на основе полученных из обучающей выборки вероятностей самих событий. Сравнительный анализ экспериментальных и моделируемых данных позволил подобрать параметр α, обеспечивающий оптимальную близость рассматриваемых распределений. При анализе использовался предложенный в [12] алгоритм применения ассоциативных функций для оценки вероятности целевого события. Выходными данными в решаемой задаче являлась функция сет-регрессии множества L на K в виде условного сет-квантиля порядка h для каждой группы пациентов. Сет-регрессионный анализ данных можно использовать для прогнозирования кризисных состояний и управления риском наступления внезапной сердечной смерти. Программную реализацию задачи сет-регрессии предполагается использовать в качестве одного из модулей в экспертной системе для поддержки принятия решения специалистом в области управления здравоохранением.

ЛИТЕРАТУРА

1. *Nguyen H. T.* An Introduction to Random Sets. Chapman and Hall/CRC, 2006. 240 p.
2. *Molchanov I.* The Theory of Random Sets. N. Y.: Springer, 2011. 488 p.
3. *Семенова Д. В., Лукьянова Н. А.* Рекуррентное построение дискретных вероятностных распределений случайных множеств событий // Прикладная дискретная математика. 2014. № 4. С. 47–58.
4. *Lukyanova N. A. and Semenova D. V.* The study of discrete probabilistic distributions of random sets of events using associative function // Журн. СФУ. Сер. Матем. и физ. 2014. Т. 7. № 4. P. 500–514.
5. *Semenova D. V.and Lukyanova N. A.* Formation of probabilistic distributions of RSE by associative functions // ITMM 2014. CCIS. Springer International Publishing Switzerland, 2014. V. 487. P. 377–386.
6. *Alsina S., Frank M., and Schveizer B.* Associative Functions: Triangular Norms and Copulas. Singapore: World Scientific Publishing Co. Pte. Ltd., 2006. 237 p.
7. *Frank M. J.* On the simultaneous associativity of $F(x, y)$ and $x + y - F(x, y)$ // Aequationes Math. 1979. No. 19. P. 194–226.

8. *Воробьев О. Ю., Воробьев А. О.* Суммирование сет-аддитивных функций и формула обращения Мёбиуса // Доклады РАН. 2009. Т. 336. № 4. С. 417–420.
9. *Воробьев О. Ю.* Сет-суммирование. Новосибирск: ВО «Наука». Сибирская издательская фирма, 1993. 137 с.
10. *Воробьев О. Ю.* Эвентология. Красноярск: СФУ, 2007. 433 с.
11. *Воробьёв О. Ю., Фомин А. Ю.* Регрессионный сет-анализ случайных событий. Красноярск: Краснояр. ун-т, 2004. 116 с.
12. *Лукьянова Н. А., Семенова Д. В.* Применение ассоциативных функций для оценки вероятности целевого события // Материалы Респуб. науч.-практич. конф. «Статистика и её применение» (Ташкент, 16–17 октября 2015 г.). Ташкент: Изд-во НУУз, 2015. С. 91–98.

REFERENCES

1. *Nguyen H. T.* An Introduction to Random Sets. Chapman and Hall/CRC, 2006. 240 p.
2. *Molchanov I.* The Theory of Random Sets. N. Y., Springer, 2011. 488 p.
3. *Semenova D. V., Lukyanova N. A.* Rekurrentnoe postroenie diskretnykh veroyatnostnykh raspredeleniy sluchaynykh mnozhestv sobytiy [Recurrent formation of discrete probabilistic distributions of random sets of events]. Prikladnaya Diskretnaya Matematika, 2014, no. 4, pp. 47–58. (in Russian)
4. *Lukyanova N. A. and Semenova D. V.* The study of discrete probabilistic distributions of random sets of events using associative function. J. SFU, Mathematics and Physics, 2014, vol. 7, no. 4, pp. 500–514.
5. *Semenova D. V.and Lukyanova N. A.* Formation of probabilistic distributions of RSE by associative functions. ITMM 2014, CCIS, Springer International Publishing Switzerland, 2014, vol. 487, pp. 377–386.
6. *Alsina S., Frank M., and Schveizer B.* Associative functions: Triangular Norms and Copulas. Singapore, World Scientific Publishing Co. Pte. Ltd., 2006. 237 p.
7. *Frank M. J.* On the simultaneous associativity of $F(x, y)$ and $x + y - F(x, y)$. Aequationes Math., 1979, no. 19, pp. 194–226.
8. *Vorob'ev O. Yu. and Vorob'ev A. O.* Summirovanie set-additivnykh funktsiy i formula obrashcheniya Mëbiusa [Summation of set-additive functions and Möbius inversion formula]. Doklady Akademii Nauk, 2009, vol. 336, no. 4, pp. 417–420. (in Russian)
9. *Vorob'ev O. Yu.* Set-summirovanie [Set-summation]. Novosibirsk, Nauka Publ., 1993. (in Russian)
10. *Vorob'ev O. Yu.* Eventologiya [Eventology]. Krasnoyarsk, SFU Publ., 2007. (in Russian)
11. *Vorob'ev O. Yu. and Fomin A. Yu.* Regressionnyy set-analiz sluchaynykh sobytiy [Regression analysis of random events]. Krasnoyarsk, KSU Publ., 2004. (in Russian)
12. *Lukyanova N. A. and Semenova D. V.* Primenenie assotsiativnykh funktsiy dlya otsenki veroyatnosti tselevogo sobytiya [The use of associative functions to estimate the target event probability]. Proc. konf. «Statistika i ee primenenie» (Tashkent, 16–17 October, 2015). Tashkent, NUUz Publ., 2015, pp. 91–98. (in Russian)

УДК 512.13

ПОСТРОЕНИЕ ПОДСТАНОВОК НА ОСНОВЕ ПОРОГОВЫХ ФУНКЦИЙ МНОГОЗНАЧНОЙ ЛОГИКИ

Д. А. Сошин

ФГУП «НИИ «Квант», г. Москва, Россия

Предложен алгоритм построения биективных отображений с помощью координатных пороговых функций k-значной логики. Алгоритм включает геометрический способ построения сбалансированных пороговых функций и два подхода к синтезу регулярных систем с приведением экспериментальных результатов.

Ключевые слова: *пороговые функции, многозначная логика, сбалансированные функции, регулярные системы.*

DOI 10.17223/20710410/32/2

CONSTRUCTING SUBSTITUTIONS ON THE BASIS OF THRESHOLD FUNCTIONS OF MULTIVALUED LOGIC

D. A. Soshin

Technology Federal State Unitary Enterprise "Research Institute Kvant", Moscow, Russia

E-mail: danil_re@list.ru

An algorithm for building one-to-one mappings (substitutions) with the help of coordinate threshold k-valued logic functions is presented. The algorithm includes a geometric way of generating balanced threshold functions and two ways to produce substitutions from these functions — by forming triangular systems and by algorithmic searching. Results of experimental testing the algorithms are given.

Keywords: *threshold functions, multiple-valued logic, balanced functions, regular systems.*

Введение

Работа посвящена построению подстановок на основе пороговых функций k-значной логики [1 – 4]. Для этого на первом этапе, в соответствии с требованиями критерия Хаффмана, строится класс сбалансированных пороговых k-значных функций от n переменных, для любых $n \geqslant 3$ и $k \geqslant 2$. Сбалансированные функции представляют и самостоятельный интерес при синтезе узлов переработки дискретной информации с точки зрения своих статистических свойств [5]. На втором этапе при помощи сбалансированных функций описанного класса строятся компактно реализуемые подстановки [4, 6]. В работе предложены два способа построения подстановок. Первый основан на формировании треугольных систем с использованием пороговых функций, которые заведомо порождают подстановки. Второй способ применяет алгоритмический поиск подстановки на базе одной из сбалансированных функций с помощью легко реализуемых операций, обобщающих преобразования движения или вращения. Доказаны утверждения, позволяющие сократить количество проверяемых вариантов. В заключение приведены результаты применения построенного алгоритма поиска регулярных систем.

1. Геометрический метод синтеза k-значных сбалансированных пороговых функций

Будем обозначать $\Omega_k = \{0, 1, \ldots, k-1\}$.

Определение 1. Функция k-значной логики $f : \Omega_k^n \longrightarrow \Omega_k$ называется пороговой, если для нее существует линейная форма

$$L(x_1, x_2, \ldots, x_n) = c_1x_1 + c_2x_2 + \ldots + c_nx_n, \quad c_i \in \mathbb{Z},\ i = 1, \ldots, n,$$

и пороги $b_0, b_1, \ldots, b_k \in \mathbb{Z}$, такие, что для любого $\alpha \in \{0, \ldots, k-1\}$ выполняется

$$f(x_1, x_2, \ldots, x_n) = \alpha \quad \Leftrightarrow \quad b_\alpha \leqslant c_1x_1 + c_2x_2 + \ldots + c_nx_n < b_{\alpha+1}.$$

Определение 2. Слоем $D_\alpha(f)$ (носителем значения α) пороговой функции f будем называть те и только те точки множества Ω_k^n, в которых функция f принимает значение α, $\alpha = 0, \ldots, k-1$. Если из контекста понятно, слой какой функции рассматривается, будем опускать символ функции и писать D_α.

В геометрическом смысле каждое неравенство $c_1x_1 + c_2x_2 + \ldots + c_nx_n < b_\alpha$ задаёт полупространство, лежащее по одну сторону от гиперплоскости L_α, определяемой равенством $c_1x_1 + c_2x_2 + \ldots + c_nx_n = b_\alpha$, а слой D_α — множество целочисленных точек n-мерного куба со стороной длины $k-1$, расположенных между двумя соседними гиперплоскостями L_α и $L_{\alpha+1}$, включая точки гиперплоскости L_α.

Далее пороги b_0 и b_k и соответствующие им гиперплоскости L_0 и L_k рассматривать не будем, поскольку они не несут смысловой нагрузки.

Функция f является сбалансированной, если слои $D_\alpha(f)$, $\alpha = 0, \ldots, k-1$, равномощные.

Для описания метода построения сбалансированных пороговых функций введём понятие среза S_α — множества точек $\{(a_1, a_2, \ldots, a_{n-1}, \alpha) \in \Omega_k^n\}$. Будем говорить, что гиперплоскость $L_\alpha(x_1, x_2, \ldots, x_n) = 0$, где $L_\alpha(x_1, x_2, \ldots, x_n) = c_1x_1 + c_2x_2 + \ldots + c_nx_n + b_\alpha$, пересекает срез, если существуют две точки этого среза, для которых выполняется одно из условий:

1) эти точки лежат по разные стороны от гиперплоскости [7];
2) одна точка принадлежит гиперплоскости, а для другой выполнено неравенство $L_\alpha(x_1, x_2, \ldots, x_n) < 0$.

Если $L_\alpha(x_1, x_2, \ldots, x_n) = 0$ — уравнение гиперплоскости, то будем говорить, что гиперплоскость отделяет (отсекает) точки среза S_α, если для этих точек выполнено условие

$$L_\alpha(a_1, a_2, \ldots, a_{n-1}, \alpha) \geqslant 0.$$

Геометрический метод построения сбалансированных пороговых функций основан на задании семейства параллельных гиперплоскостей L_α, $\alpha = 1, \ldots, k-1$, в n-мерном пространстве, каждая гиперплоскость которого проходит через соответствующий набор точек:

$$\begin{gathered} t^{(1)} = (r, 0, 0, 0, \ldots, 0, 0, \alpha), \\ t^{(2)} = (0, r, 0, 0, \ldots, 0, 0, \alpha), \\ \ldots \\ t^{(n-1)} = (0, 0, 0, 0, \ldots, 0, r, \alpha), \\ t^{(n)} = (k-r, k-1, \ldots, k-1, \alpha-1), \end{gathered} \tag{1}$$

то есть при подстановке их в $L_\alpha(x_1, x_2, \ldots, x_n)$ выполняется равенство

$$L_\alpha\left(t^{(j)}\right) = 0, \quad j = 1, \ldots, n.$$

Проходя через точки $t^{(1)}, t^{(2)}, \ldots, t^{(n-1)}$, гиперплоскость L_α отделяет в срезе S_α множество точек

$$\left\{(a_1, a_2, \ldots, a_{n-1}, \alpha) \in \Omega_k^n : a_1 + \ldots + a_{n-1} \geqslant r\right\},$$

которые не войдут в слой $D_{\alpha-1}$. Прохождение гиперплоскости через точку $t^{(n)}$ позволяет отделить в срезе $S_{\alpha-1}$ множество точек

$$\left\{(a_1, a_2, \ldots, a_{n-1}, \alpha - 1) \in \Omega_k^n : a_1 + \ldots + a_{n-1} \geqslant (k-1)(n-1) - r + 1\right\},$$

которые войдут в слой $D_{\alpha-1}$. В силу равномощности множеств

$$\left\{(a_1, a_2, \ldots, a_{n-1}, \alpha) \in \Omega_k^n : a_1 + \ldots + a_{n-1} \leqslant r - 1\right\}$$
$$\text{и} \quad \left\{(a_1, a_2, \ldots, a_{n-1}, \alpha - 1) \in \Omega_k^n : a_1 + \ldots + a_{n-1} \geqslant (k-1)(n-1) - r + 1\right\}$$

и в случае, если гиперплоскости L_α не пересекают более двух срезов, такая последовательная компенсация для каждого слоя позволяет сохранит сбалансированность.

Расстояние r будем называть *отступом* точек $t^{(1)}, t^{(2)}, \ldots, t^{(n-1)}$ от точки $(0, 0, \ldots, 0, \alpha)$ среза S_α. Отступ первых $n-1$ точек в системе (1) связан с удобством задания точки $t^{(n)}$ в срезе $S_{\alpha-1}$, через которую пройдёт гиперплоскость. Для случая произвольного выбора отступов необходимо найти самую близкую целочисленную точку среза $S_{\alpha-1}$ к прямой (для $n \geqslant 4$ — к гиперплоскости) пересечения плоскости L_α и плоскости $\widetilde{S}_{\alpha-1}$, описываемой уравнением $x_n = \alpha - 1$. В рассмотренном случае выбор последней точки не вызывает сложности. Для произвольных параметров n, k и произвольных отступов от крайней точки последняя задача не имеет общего решения, но для каждого фиксированного случая её можно решить и построить любую сбалансированную пороговую функцию.

Следующая теорема описывает пороговые сбалансированные функции, соответствующие данному семейству гиперплоскостей, задаваемых системами точек (1).

Теорема 1. Пусть $R_n = k - 2r + (n-2)(k-1)$, $P_\alpha^n = r + \alpha R_n$ для некоторых $k \geqslant 2$, $r \geqslant 1$, $n \geqslant 2$; $\alpha = 1, \ldots, k-1$. Пороговая функция $f(x_1, x_2, \ldots, x_n)$, заданная двусторонними неравенствами

$$f(x_1, x_2, \ldots, x_n) = \alpha \quad \Leftrightarrow \quad b_\alpha \leqslant x_1 + x_2 + \ldots + R_n x_n < b_{\alpha+1}, \tag{2}$$

где $(b_1, b_2, \ldots, b_{k-1}) = (P_1^n, P_2^n, \ldots, P_{k-1}^n)$, при $(k-1)(n-1) > 3r - 2$ является сбалансированной пороговой.

Доказательство. Покажем, что пороговая функция (2) соответствует семейству гиперплоскостей L_α, отвечающим системам точек (1). Пусть уравнения семейства гиперплоскостей L_α имеют вид

$$c_1 x_1 + c_2 x_2 + \ldots + c_n x_n = b_\alpha.$$

Покажем, что значения порогов и коэффициентов линейной формы в условии теоремы равны соответствующим параметрам:

$$(b_1, b_2, \ldots, b_{k-1}) = (P_1^n, P_2^n, \ldots, P_{k-1}^n), \quad (c_1, c_2, \ldots, c_n) = (1, 1, \ldots, 1, R_n).$$

Уравнение $(n-1)$-мерной гиперплоскости, проходящей через точки $t^{(1)}, t^{(2)}, \ldots, t^{(n)}$, которые не лежат на одной $(n-2)$-мерной гиперплоскости, задаётся следующим образом [7]:

$$\left|\begin{pmatrix} x_1 - t_1^{(1)} & x_2 - t_2^{(1)} & \ldots & x_{n-1} - t_{n-1}^{(1)} & x_n - t_n^{(1)} \\ t_1^{(2)} - t_1^{(1)} & t_2^{(2)} - t_2^{(1)} & \ldots & t_{n-1}^{(2)} - t_{n-1}^{(1)} & t_n^{(2)} - t_n^{(1)} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ t_1^{(n-1)} - t_1^{(1)} & t_2^{(n-1)} - t_2^{(1)} & \ldots & t_{n-1}^{(n-1)} - t_{n-1}^{(1)} & t_n^{(n-1)} - t_n^{(1)} \\ t_1^{(n)} - t_1^{(1)} & t_2^{(n)} - t_2^{(1)} & \ldots & t_{n-1}^{(n)} - t_{n-1}^{(1)} & t_n^{(n)} - t_n^{(1)} \end{pmatrix}\right| = 0.$$

Найдём данный определитель для рассматриваемого случая:

$$\Delta = \left|\begin{pmatrix} x_1 - r & x_2 & x_3 & x_4 & \ldots & x_{n-1} & x_n - \alpha \\ -r & r & 0 & 0 & \ldots & 0 & 0 \\ -r & 0 & r & 0 & \ldots & 0 & 0 \\ -r & 0 & 0 & r & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ -r & 0 & 0 & 0 & \ldots & r & 0 \\ k-2r & k-1 & k-1 & k-1 & \ldots & k-1 & -1 \end{pmatrix}\right|.$$

Добавим к первому столбцу все столбцы, кроме последнего. Столбец с номером n, умноженный на $k - 2r + (n-2)(k-1)$, прибавим к первому столбцу; к остальным столбцам добавим его же, умноженным на $k-1$. Через $*$ обозначим элементы матрицы, которые не влияют на значение определителя. В результате получим

$$\Delta = \left|\begin{pmatrix} x_1 + x_2 + \ldots + x_{n-1} + (x_n - \alpha)(k - 2r + (n-2)(k-1)) - r & * & * & \ldots & * & * \\ 0 & r & 0 & \ldots & 0 & 0 \\ 0 & 0 & r & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \ldots & r & 0 \\ 0 & 0 & 0 & \ldots & 0 & -1 \end{pmatrix}\right| =$$

$$= -r^{n-2}\left(x_1 + x_2 + \ldots + x_{n-1} + (x_n - \alpha)(k - 2r + (n-2)(k-1)) - r\right).$$

Приравняем $\Delta = 0$ и перенесём свободный член в правую часть:

$$-r^{n-2}(x_1 + x_2 + \ldots + x_{n-1} + (x_n - \alpha)(k - 2r + (n-2)(k-1)) - r) = 0,$$
$$x_1 + x_2 + \ldots + x_{n-1} + (x_n - \alpha)(k - 2r + (n-2)(k-1)) - r = 0,$$
$$x_1 + x_2 + \ldots + x_{n-1} + x_n(k - 2r + (n-2)(k-1)) = \alpha(k - 2r + (n-2)(k-1)) + r,$$
$$x_1 + x_2 + \ldots + x_{n-1} + x_n R_n = P_\alpha^n.$$

Последнее равенство доказывает требуемое соответствие.

Докажем выполнение условия сбалансированности. Для этого необходимо проверить, что гиперплоскость L_α не пересекает срез $S_{\alpha-2}$. Для выполнения этого условия достаточно, чтобы линейная форма $L = x_1 + \ldots + x_{n-1} + x_n R_n$ в точке $t = (k-1, k-1, \ldots, k-1, \alpha-2)$ принимала значение строго меньше P_α^n:

$$L(t) = (n-1)(k-1) + (\alpha-2)R_n < P_\alpha^n = r + \alpha R_n.$$

Последовательно получаем

$$r + \alpha R_n > (n-1)(k-1) + (\alpha - 2)R_n,$$
$$r > (n-1)(k-1) - 2R_n.$$
$$r > (k-1)(n-1) - 2(k - 2r + (n-2)(k-1)),$$
$$r > -(k-1)(n-1) + 4r - 2,$$
$$(k-1)(n-1) > 3r - 2.$$

Последнее неравенство выполнено по условию. ■

Замечание 1. При добавлении фиктивных переменных у функций, построенных в теореме 1, свойство сбалансированности не нарушается. Теорема 2 описывает сбалансированные функции построенного типа с произвольным количеством фиктивных переменных.

Теорема 2. В обозначениях теоремы 1 при $(k-1)(n-m-1) > 3r-2, 0 \leqslant m \leqslant n-2$ пороговая функция $f^k(x_1, x_2, \ldots, x_n)$, заданная двусторонними неравенствами

$$f(x_1, x_2, \ldots, x_n) = \alpha \quad \Leftrightarrow \quad b_\alpha \leqslant x_{m+1} + x_{m+2} + \ldots + x_n R_{n-m} < b_{\alpha+1}, \tag{3}$$

где $(b_1, b_2, \ldots, b_{k-1}) = (P_1^{n-m}, P_2^{n-m}, \ldots, P_{k-1}^{n-m})$, является сбалансированной пороговой.

Доказательство. Справедливость теоремы 2 следует из замечания 1 и теоремы 1 при замене n на $n - m$. ■

При $m = 0$ утверждение теоремы 2 совпадает с утверждением теоремы 1.

Приведём пример построения функции для фиксированных параметров с неодинаковыми отступами.

Пример 1. Пусть $n = 3$, $k = 3$, отступы от крайней точки равны 2 по первой и 1 по второй переменной. Тогда набор точек, через которые пройдёт соответствующее семейство плоскостей, следующий:

$$t^{(1)} = (2, 0, \alpha), \quad t^{(2)} = (0, 1, \alpha), \quad t^{(3)} = (1, 2, \alpha - 1), \quad \alpha \in \{1, 2\}.$$

Найдём уравнения плоскостей, проходящих через точки $t^{(1)}, t^{(2)}, t^{(3)}$; для этого необходимо найти соответствующий определитель Δ и приравнять его к нулю:

$$\Delta = \left| \begin{pmatrix} x_1 - 2 & x_2 & x_3 - \alpha \\ -2 & 1 & 0 \\ -1 & 2 & -1 \end{pmatrix} \right| = \left| \begin{pmatrix} x_1 - x_3 + \alpha - 2 & x_2 + 2x_3 - 2\alpha & x_3 - \alpha \\ -2 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix} \right| =$$
$$= \left| \begin{pmatrix} x_1 - x_3 + \alpha - 2 + 2(x_2 + 2x_3 - 2\alpha) & x_2 + 2x_3 - 2\alpha & x_3 - \alpha \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix} \right| =$$
$$= -(x_1 - x_3 + \alpha - 2 + 2(x_2 + 2x_3 - 2\alpha)) = -(x_1 + 2x_2 + 3x_3 - 3\alpha - 2).$$

Приравнивая $\Delta = 0$, получим уравнения плоскостей для $\alpha \in \{1, 2\}$:

$$x_1 + 2x_2 + 3x_3 = 3\alpha + 2.$$

Вектор порогов данной функции следующий: $(b_1, b_2) = (5, 8)$. Прохождение плоскостей через соответствующий куб изображено на рис. 1.

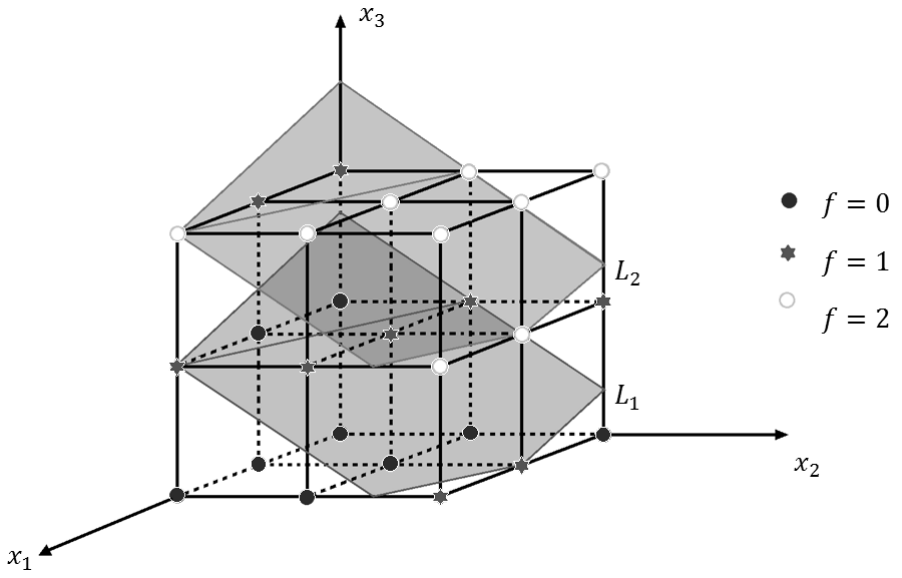

Рис. 1. Функция, построенная в примере 1 на трёхмерном кубе трёхзначной логики; L_1 и L_2 — построенные плоскости

2. Построение подстановок на основе сбалансированных k-значных пороговых функций

Рассмотрим два способа построения подстановок. Система координатных функций задаёт подстановку, если она регулярна. Проверка системы на регулярность является сложной задачей. Существующие критерии, в частности критерий Хаффмана, не позволяют оптимизировать проверку на регулярность, но в некоторых случаях дают возможность сузить множество проверяемых на регулярность систем, отбраковав заведомо ложные.

Определение 3. Координатное отображение $F : \Omega_k^n \to \Omega_k^n$,

$$F(x_1, x_2, \ldots, x_n) = (f_1(x), f_2(x), \ldots, f_n(x)), \tag{4}$$

где $f_i(x) = f_i(x_1, x_2, \ldots, x_n)$ — k-значные функции, называется биективным, а система функций $f_1(x), f_2(x), \ldots, f_n(x)$ — регулярной системой, если F — взаимно-однозначное отображение.

Первый предлагаемый способ основан на построении треугольных регулярных систем, второй — на построении подстановок, координатные функции которых являются однотипными k-значными функциями, полученными из одной путём действия преобразований перестановки переменных и их инвертирования.

2.1. Построение подстановок на основе сбалансированных k-значных пороговых функций

Определение 4. Система функций $f_1(x), f_2(x), \ldots, f_n(x)$ называется треугольной, если она имеет вид

$$\begin{cases} f_1(x) = \varphi_1(x_1), \\ f_2(x) = \varphi_2(x_1, x_2), \\ \ldots \\ f_i(x) = \varphi_i(x_1, x_2, \ldots, x_i), \\ \ldots \\ f_n(x) = \varphi_n(x_1, x_2, \ldots, x_n), \end{cases} \tag{5}$$

где $\varphi_1, \varphi_2, \ldots, \varphi_n$ — произвольные k-значные функции.

Если функция φ_i, $i = 1, \ldots, n$, биективна по переменной x_i, т. е. при каждой фиксации переменных $x_1, x_2, \ldots, x_{i-1}$ функция φ_i задаёт подстановочное отображение [8, с. 166], то система (5) задаёт биекцию. Действительно, пусть задано равенство

$$(\varphi_1, \varphi_2, \ldots, \varphi_n)(x_1, x_2, \ldots, x_n) = (a_1, a_2, \ldots, a_n), \tag{6}$$

тогда $\varphi_1(x_1) = (a_1)$. В силу биективности функции $\varphi_1(x_1)$ по переменной x_1 существует и единственен элемент $b_1 \in \Omega_k$, такой, что $\varphi_1(b_1) = a_1$. Положим $x_1 = b_1$. Ввиду биективности функции φ_2 по переменной x_2 существует и единственен $b_2 \in \Omega_k$, такой, что $\varphi_2(b_1, b_2) = a_2$. Положим $x_2 = b_2$. Продолжая дальше по индукции, получаем, что существует единственный набор $(x_1, x_2, \ldots, x_n) = (b_1, b_2, \ldots, b_n) \in \Omega_k^n$, удовлетворяющий равенству (6).

В качестве треугольной регулярной системы можно предложить систему вида

$$\begin{cases} \varphi_1(x_1) = x_1, \\ \varphi_2(x_1, x_2) = x_2 + \psi_2(x_1) \mod k, \\ \ldots \\ \varphi_n(x_1, x_2, \ldots, x_n) = x_n + \psi_n(x_1, x_2, \ldots, x_{n-1}) \mod k. \end{cases} \tag{7}$$

Данная система регулярна для любых k-значных функций $\psi_2, \psi_3, \ldots, \psi_n$, поскольку каждая φ_i биективна по крайней переменной независимо от значений функции ψ_i.

Среди систем (7) представляют интерес системы на основе функций (3) из теоремы 2. Пороговые функции (3) будем обозначать следующим образом:

$$T_r^m(x_1, x_2, \ldots, x_n), \tag{8}$$

где m — количество фиктивных переменных; r — параметр, отвечающий за размер отступа от крайних точек среза, тем самым отвечающий за размер отсекаемой области среза. Систему порогов и максимальный коэффициент линейной формы будем обозначать $(b_1, b_2, \ldots, b_{k-1}) = \left(P_1^{n-m}(r), P_2^{n-m}(r), \ldots, P_{k-1}^{n-m}(r)\right)$ и $R_{n-m}(r)$ соответственно.

Для дальнейшего использования доопределим множество функций (8) для случая $n = 1$: $T_r^0(x_1) = x_1$, $R_1(r) = 1$, $P_i(r) = i$, $i = 1, \ldots, k-1$.

Можно рассмотреть два способа задания функций $\psi_2, \psi_3, \ldots, \psi_n$. Первый способ позволяет параллельно вычислять значения координатных функций, что ускоряет ре-

ализацию подстановки, а именно:

$$\begin{cases} \psi_2(x_1) = T_{r_1}^{n-1}(0,0,\ldots,0,x_1), \\ \psi_3(x_1,x_2) = T_{r_2}^{n-2}(0,0,\ldots,0,x_1,x_2), \\ \ldots \\ \psi_i(x_1,x_2,\ldots,x_{i-1}) = T_{r_{i-1}}^{n-i+1}(0,0,\ldots,0,x_1,x_2,\ldots,x_{i-1}), \\ \ldots \\ \psi_n(x_1,x_2,\ldots,x_{n-1}) = T_{r_{n-1}}^{1}(0,x_1,x_2,\ldots,x_{n-1}). \end{cases} \tag{9}$$

Второй способ реализует рекуррентное определение значений координатных функций, основанное на последовательном задании функций ψ_i:

$$\begin{cases} \psi_2(y_1) = T_{r_1}^{n-1}(0,0,\ldots,0,y_1), \\ \psi_3(y_1,y_2) = T_{r_2}^{n-2}(0,0,\ldots,0,y_1,y_2), \\ \ldots \\ \psi_i(y_1,y_2,\ldots,y_{i-1}) = T_{r_{i-1}}^{n-i+1}(0,0,\ldots,0,y_1,y_2,\ldots,y_{i-1}), \\ \ldots \\ \psi_n(y_1,y_2,\ldots,y_{n-1}) = T_{r_{n-1}}^{1}(0,y_1,y_2,\ldots,y_{n-1}), \end{cases} \tag{10}$$

где

$$\begin{cases} y_1(x_1) = \varphi_1(x_1) = x_1, \\ y_2(x_1,x_2) = \varphi_2(x_1,x_2) = x_2 + \psi_2(x_1) \mod k, \\ \ldots \\ y_{n-1}(x_1,x_2,\ldots,x_{n-1}) = \varphi_{n-1}(x_1,x_2,\ldots,x_{n-1}) = x_{n-1} + \psi_n(x_1,x_2,\ldots,x_{n-2}) \mod k. \end{cases}$$

Удобство использования систем (7) с функциями (9) или (10) заключается в том, что для задания подстановки достаточно хранить коэффициенты линейных форм пороговых функций и систему порогов. Матрицы C и P содержат всю информацию о задаваемых подстановках указанного типа:

$$C = \begin{pmatrix} 0 & 0 & 0 & \ldots & 0 & 0 & R_1(r_1) \\ 0 & 0 & 0 & \ldots & 0 & 1 & R_2(r_2) \\ 0 & 0 & 0 & \ldots & 1 & 1 & R_3(r_3) \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 1 & \ldots & 1 & 1 & R_{n-2}(r_{n-2}) \\ 0 & 1 & 1 & \ldots & 1 & 1 & R_{n-1}(r_{n-1}) \end{pmatrix}, \tag{11}$$

$$P = \begin{pmatrix} P_1^1(r_1) & P_2^1(r_1) & \ldots & P_{k-1}^1(r_1) \\ P_1^2(r_2) & P_2^2(r_2) & \ldots & P_{k-1}^2(r_2) \\ \vdots & \vdots & \ddots & \vdots \\ P_1^{n-1}(r_{n-1}) & P_2^{n-1}(r_{n-1}) & \ldots & P_{k-1}^{n-1}(r_{n-1}) \end{pmatrix}.$$

Элемент $c_{i,j}$ матрицы C задаёт коэффициент линейной формы i-й координатной пороговой функции при переменной x_j. Например, линейная форма i-й пороговой функции имеет вид $L = x_1 + x_2 + \ldots + x_{i-1} + x_i R_i(r_i)$ и соответствует линейной форме функции $\psi_i(x_1,x_2,\ldots,x_i)$ систем (9) и (10). В матрице P строка с номером i задаёт систему порогов i-й пороговой функции.

2.2. Построение подстановок на основе систем однотипных пороговых функций

Второй способ основан на построении подстановок, координатные функции которых являются однотипными k-значными функциями, полученными из одной (стрелки Лукашевича) путём действия преобразований перестановки переменных и их инвертирования. В этом способе представления нашли своё воплощение три идеи, каждая из которых определяет потенциальные преимущества при реализации подстановки.

Во-первых, в предложенном способе заложен принцип компактной реализации, при котором для вычисления координат результирующего вектора используется единая функция с простым сервисным преобразованием, аналогичным преобразованию однотипности в булевом случае.

Во-вторых, по всем координатам функции, задающие отображение, — пороговые, для которых в перспективной элементной базе, например оптической, может быть осуществлена реализация в среде-носителе сигнала с высоким быстродействием.

И наконец, третьим преимуществом является сбалансированность и временнàя синхронизация выработки значений всех координат выходного вектора.

Остановимся подробно на преобразованиях, используемых при генерации функций по каждому каналу. Группой движения G_n назовем группу, порождённую группами S_n и N_n:

$$G_n = \langle S_n, N_n \rangle,$$

где S_n — группа подстановок на множестве $\{1, \dots, n\}$; $N_n = \{-1, 1\}^n$ — группа инвертирования переменных. Действие данных групп на множестве Ω_k^n определяется следующим образом: для любой точки $(a_1, a_2, \dots, a_n) \in \Omega_k^n$, для любых преобразований $s \in S_n$ и $\beta = (\beta_1, \beta_2, \dots, \beta_n) \in N_n$

$$(a_1, a_2, \dots, a_n)^s = (a_{s^{-1}(1)}, a_{s^{-1}(2)}, \dots, a_{s^{-1}(n)}),$$

$$(a_1, a_2, \dots, a_n)^\beta = (u_1, u_2, \dots, u_n), \qquad u_i = \begin{cases} a_i, \text{ если } \beta_i = 1, \\ k - 1 - a_i, \text{ если } \beta_i = -1. \end{cases}$$

Обозначим через $(\beta s)\, T_r^m$ функцию, полученную из функции T_r^m по правилу $(\beta s)\, T_r^m = T_r^m(x^{\beta s})$. Системе k-значных пороговых функций

$$F_r^m(\beta, s) = \left(\beta^{(1)} s^{(1)}\right) T_r^m, \left(\beta^{(2)} s^{(2)}\right) T_r^m, \dots, \left(\beta^{(n)} s^{(n)}\right) T_r^m, \tag{12}$$

где $\beta = (\beta^{(1)}, \beta^{(2)}, \dots, \beta^{(n)}) \in (N_n)^n$; $s = (s^{(1)}, s^{(2)}, \dots, s^{(n)}) \in (S_n)^n$, поставим в соответствие матрицу $C = (c_{i,j})_{n \times n}$ аналогично матрице (11) с учётом действия преобразований из группы N_n. Например, функции $(\beta\varepsilon) T_r^m$, где ε — нейтральный элемент группы S_n, $\beta = (\beta_1, \beta_2, \dots, \beta_n)$, соответствует строка матрицы $(\beta_1, \beta_2, \dots, \beta_n R_{n-m}(r))$. Каждой системе (12) соответствует отображение, задаваемое по правилу

$$\begin{gathered} \pi[F_r^m(\beta, s)](x_1, x_2, \dots, x_n) = \\ = \left(\left(\beta^{(1)} s^{(1)}\right) T_r^m(x_1,\ x_2, \dots, x_n), \dots, \left(\beta^{(n)} s^{(n)}\right) T_r^m(x_1, x_2, \dots, x_n)\right). \end{gathered} \tag{13}$$

В случае регулярности системы $F_r^m(\beta, s)$ отображение $\pi[F_r^m(\beta, s)]$ задаёт подстановку на множестве Ω_k^n.

Поиск регулярной системы осуществляется алгоритмическим способом, а именно путём перебора функций $\left(\beta^{(i)} s^{(i)}\right) T_r^m$ и проверки факта порождения подстановки. Приведём формулировку критерия Хаффмана и утверждение, позволяющее сократить количество проверяемых систем $F_r^m(\beta, s)$ на регулярность для произвольных параметров m и r, удовлетворяющих условиям теоремы 2.

Теорема 3 (критерий Хаффмана). Система k-значных функций $f_i(x_1, \ldots, x_n)$, $i = 1, \ldots, n$, регулярна тогда и только тогда, когда для любых $r \in \{1, \ldots, n\}$, $1 \leqslant i_1 < i_2 < \ldots < i_r \leqslant n$ и $\alpha_1, \alpha_2, \ldots, \alpha_r \in \{0, \ldots, k-1\}$ выполняется

$$\left| \bigcap_{w=1}^{r} D_{\alpha_w}(f_{i_w}) \right| = k^{n-r}.$$

Теорема 4. Пусть для некоторых $(s^{(1)}, s^{(2)}, \ldots, s^{(n)}) \in (S_n)^n$ и любых $\beta = (\beta^{(1)}, \beta^{(2)}, \ldots, \beta^{(n)}) \in (N_n)^n$ при $k > 3$ в матрице C, отвечающей системе $F_r^m(\beta, s)$, существуют две различные строки вида

$$\left(\theta_1^i\beta_1^{(i)} \quad \theta_2^i\beta_2^{(i)} \quad \ldots \quad \theta_{p-1}^i\beta_{p-1}^{(i)} \quad \beta_p^{(i)}R_{n-m}(r) \quad \theta_{p+1}^i\beta_{p+1}^{(i)} \quad \ldots \quad \theta_n^i\beta_n^{(i)}\right),$$
$$\left(\theta_1^j\beta_1^{(j)} \quad \theta_2^j\beta_2^{(j)} \quad \ldots \quad \theta_{p-1}^j\beta_{p-1}^{(j)} \quad \beta_p^{(j)}R_{n-m}(r) \quad \theta_{p+1}^j\beta_{p+1}^{(j)} \quad \ldots \quad \theta_n^j\beta_n^{(j)}\right),$$

где θ_u^i, θ_u^j — индикаторы фиктивности (существенности) соответствующих переменных; $\sum\limits_{u \neq p} \theta_u^i = \sum\limits_{u \neq p} \theta_u^j = n - m + 1$. Тогда отображение $\pi[F_r^m(\beta, s)]$ не является биекцией.

Доказательство. Не ограничивая общности, положим $p = n$. Пусть f и g — функции, соответствующие строкам i и j матрицы C. Зафиксируем $\alpha = \beta_n^{(i)}(0)$ и $\delta = \beta_n^{(j)}(k-1)$. Тогда справедливы включения $D_\alpha(f) \subset S_0 \cup S_1$, $D_\delta(g) \subset S_{k-2} \cup S_{k-1}$. Поскольку $k > 3$, то $(S_0 \cup S_1) \cap (S_{k-2} \cup S_{k-1}) = \varnothing$. Из последних трёх выражений получаем $D_\alpha(f) \cap D_\delta(g) = \varnothing$, что противоречит условию критерия Хаффмана. ■

Следствие 1. Теорема 4 позволяет сократить перебор отображений (13), проверяемых на регулярность при фиксированных параметрах r и m, где $(k-1)(n-m-1) > 3r - 2$, в n^n раз

Доказательство. Действительно, тотальный поиск биективных отображений (13) состоит из задания соответствующей матрицы C и проверки на регулярность порождаемого ею преобразования. Выбор матриц C осуществляется за счёт расположения коэффициента $R_{n-m}(r)$ в каждой строке (n^n вариантов), задания у каждой координатной функции фиктивных переменных (C_{n-1}^m вариантов) и определения инвертирования существенных переменных, что потребует рассмотрения $2^{(n-m)n}$ вариантов, где $(n-m)n$ — количество ненулевых элементов матрицы C. Следовательно, на основе матрицы C проверке подвергаются $N = n^n C_{n-1}^m 2^{(n-m)n}$ систем.

Из теоремы 4 следует, что в каждом столбце и в каждой строке матрицы, соответствующей отображению (13), должен присутствовать единственный максимальный по модулю коэффициент $\beta_p^{(j)}R_{n-m}(r)$. Перестановкой строк матрицы C можно расположить коэффициенты $\beta_p^{(j)}R_{n-m}(r)$ на главной диагонали, поскольку перестановка координатных функций не нарушает биективности. Таким образом, перебор заключается в фиксации фиктивных переменных и расстановке инвертирования существенных переменных, что составляет $C_{n-1}^m 2^{(n-m)n}$ вариантов. ■

Приведём алгоритм поиска биективных отображений (алгоритм 1).

Приведём результаты применения программы, реализующей данный алгоритм на основе матрицы C. Во всех случаях n — размерность пространства, k — значность логики, m — количество фиктивных переменных, r — параметр, отвечающий за размер отсекаемой области, $R_{n-m}(r)$ определено условием теоремы 2, $\beta^{(1)}, \beta^{(2)}, \ldots, \beta^{(n)} \in N_n$ — перебираемые параметры.

Алгоритм 1. Поиск биективных отображений

1: Инициализируем параметры $j := 0$, n, k, m, r, $R_{n-m}(r)$, $P_0^{n-m}(r)$, $(P_1^{n-m}(r), \ldots, P_{k-2}^{n-m}(r))$.
2: Проверяем условие сбалансированности $(k-1)(n-m-1) > 3r+2$. Если выполнено, то на шаг 3, иначе сообщение «Функция не сбалансирована», конец алгоритма.
3: Инициализируем матрицу коэффициентов C. На главной диагонали располагаем максимальный коэффициент $R_{n-m}(r)$, оставшиеся элементы каждой строчки заполняем m нулями и $n-m-1$ единицами согласно рассматриваемому случаю.
4: Для каждого из 2^{n-m} способов задаём знаки ненулевых коэффициентов матрицы C, проверяем на биективность получившуюся систему. Если биекция, то $j := j+1$.

Выход: j — количество найденных подстановок для проверяемого случая.

С л у ч а й 1. Общие параметры $m = 0$,

$$C^{(1)} = \begin{pmatrix} \beta_1^{(1)} R_n(r) & \beta_2^{(1)} & \ldots & \beta_{n-1}^{(1)} & \beta_n^{(1)} \\ \beta_1^{(2)} & \beta_2^{(2)} R_n(r) & \ldots & \beta_{n-1}^{(2)} & \beta_n^{(2)} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \beta_1^{(n-1)} & \beta_2^{(n-1)} & \ldots & \beta_{n-1}^{(n-1)} R_n(r) & \beta_n^{(n-1)} \\ \beta_1^{(n)} & \beta_2^{(n)} & \ldots & \beta_{n-1}^{(n)} & \beta_n^{(n)} R_n(r) \end{pmatrix}.$$

Для заданных m и $C^{(1)}$ рассмотрим два способа задания размера отступа r.

Первый способ соответствует $r = 1$, количество подстановок, полученных за счёт перебора $\beta^{(1)}, \beta^{(2)}, \ldots, \beta^{(n)} \in N_n$, представлено в табл. 1.

Т а б л и ц а 1

	k														
n	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
2	–	8	8	8	8	8	8	8	8	8	8	8	8	8	8
3	192	64	64	64	64	64	64	64	64	64	64	64	64	64	64
4	3328	768	768	768	768	768	768	–	–	–	–	–	–	–	–
5	76800	12288	12288	–	–	–	–	–	–	–	–	–	–	–	–

Во втором способе для каждого варианта параметров n и k будем задавать максимально возможный размер отступа $r = r_{\max}(k, n, m)$, удовлетворяющий условию сбалансированности теоремы 2. Значения $r_{\max}(k, n, m)$ и количества подстановок, полученных за счёт перебора $\beta^{(1)}, \beta^{(2)}, \ldots, \beta^{(n)} \in N_n$, представлены в табл. 2.

Т а б л и ц а 2

	k												
n	2	3	4	5	6	7	8	9	10	11	12	13	14
3	1; 192	1; 64	2; 0	3; 0	3; 0	4; 0	5; 0	5; 0	6; 0	7; 0	7; 0	8; 0	9; 0
4	1; 3328	2; 0	3; 0	4; 0	5; 0	6; 0	7; 0	–	–	–	–	–	–
5	1; 76800	3; 0	3; 0	–	–	–	–	–	–	–	–	–	–

По результатам, представленным в табл. 1, можно сделать предположение, что с ростом k количество подстановок на основе матрицы $C^{(1)}$ стабилизируется. Из табл. 2

можно предположить, что для значений $1 < r \leqslant r_{\max}(k,n,m)$ подстановок указанного вида нет.

С л у ч а й 2. Общие параметры $m = 1$,

$$C^{(2)} = \begin{pmatrix} \beta_1^{(1)} R_{n-1}(r) & 0 & \beta_3^{(1)} & \dots & \beta_{n-1}^{(1)} & \beta_n^{(1)} \\ \beta_1^{(2)} & \beta_2^{(2)} R_{n-1}(r) & 0 & \dots & \beta_{n-1}^{(2)} & \beta_n^{(2)} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \beta_1^{(n-1)} & \beta_2^{(n-1)} & \beta_3^{(n-1)} & \dots & \beta_{n-1}^{(n-1)} R_{n-1}(r) & 0 \\ 0 & \beta_2^{(n)} & \beta_3^{(n)} & \dots & \beta_{n-1}^{(n)} & \beta_n^{(n)} R_{n-1}(r) \end{pmatrix},$$

где нули расположены на второй главной диагонали.

Для $r = 1$ нашлись биекции только для случаев $k = 2$, $n = 3, 4, 5$.

Для $r = r_{\max}(k,n,m)$ биекции были найдены только при $k = 2$, $n = 3, 4, 5$, при этом $r_{\max}(k,n,m) = 1$.

При $r > 1$ биекций на основе матрицы $C^{(2)}$ не найдено.

С л у ч а й 3. Общие параметры $m = 1$, $n = 4$,

$$C^{(3)} = \begin{pmatrix} \beta_1^{(1)} R_3(r) & \beta_2^{(1)} & \beta_3^{(1)} & 0 \\ \beta_1^{(2)} & \beta_2^{(2)} R_3(r) & 0 & \beta_4^{(2)} \\ \beta_1^{(3)} & 0 & \beta_3^{(3)} R_3(r) & \beta_4^{(3)} \\ 0 & \beta_2^{(4)} & \beta_3^{(4)} & \beta_4^{(4)} R_3(r) \end{pmatrix}.$$

Для $r = 1$, $k = 1, \dots, 8$ нашлись биекции только для случая $k = 2$.

Для $r = r_{\max}(k,n,m)$, $k = 1, \dots, 8$ результат совпадает со случаем $r = 1$.

Отсюда сделаем вывод, что возможно построение подстановок с однотипными координатными пороговыми функциями из теоремы 2 и предположительно только для размера отступа $r = 1$.

Приведём утверждение, использующее блочную структуру и позволяющее итеративно увеличивать размер подстановки, основанной на пороговых функциях.

Теорема 5. Пусть заданы две подстановки π_1, π_2, основанные на пороговых функциях, такие, что

$$\pi_1 : \Omega_k^{n_1} \to \Omega_k^{n_1}, \quad \pi_2 : \Omega_k^{n_2} \to \Omega_k^{n_2}, \quad n_1, n_2 \geqslant 2;$$

матрица $C^{(v)} = \left(c_{i,j}^{(v)}\right)_{n_v \times n_v}$ — матрица коэффициентов линейных форм и $P^{(v)} = \left(p_{i,j}^{(v)}\right)_{n_v \times (k-1)}$ — матрица порогов координатных функций подстановки π_v, $v = 1, 2$; $\Theta^{(1)}, \Theta^{(2)}$ — нулевые матрицы размеров $n_1 \times n_2$ и $n_2 \times n_1$ соответственно. Тогда матрицы

$$\tilde{C} = \begin{pmatrix} C^{(1)} & \Theta^{(1)} \\ \Theta^{(2)} & C^{(2)} \end{pmatrix}, \quad \tilde{P} = \begin{pmatrix} P^{(1)} \\ P^{(2)} \end{pmatrix}$$

задают подстановку $\tilde{\pi} : \Omega_k^{n_1+n_2} \to \Omega_k^{n_1+n_2}$.

Доказательство. Достаточно заметить, что первые n_1 координатных функций фиктивно зависят от последних n_2 переменных и реализуют подстановку π_1 на первых n_1 переменных. Последние n_2 координатных функций реализуют подстановку π_2 на последних n_2 переменных. Поэтому $\tilde{\pi}$ представляется следующим образом:

$$\tilde{\pi}(x_1, x_2, \dots, x_{n_1+n_2}) = (\pi_1(x_1, x_2, \dots, x_{n_1}), \pi_2(x_{n_1+1}, x_{n_1+2}, \dots, x_{n_1+n_2})).$$

Данное представление задаёт подстановку. ■

ЛИТЕРАТУРА

1. *Никонов В. Г., Саранцев А. В.* Методы компактной реализации биективных отображений, заданных регулярными системами однотипных булевых функций // Вестник Российского университета дружбы народов. Сер. Прикладная и компьютерная математика. 2003. Т. 2. № 1. С. 94–105.
2. *Никонов В. Г., Саранцев А. В.* Построение и классификация регулярных систем однотипных функций // Материалы XXXI Междунар. конф. «Информационные технологии в науке, образовании, телекоммуникации и бизнесе». М., 2004. Т. 5. С. 173–174.
3. *Никонов В. Г., Сидоров Е. С.* О способе построения взаимно однозначных отображений при помощи квазиадамаровых матриц // Вестник Московского государственного университета леса — Лесной вестник. 2009. № 2(65). С. 155–157.
4. *Никонов В. Г., Сошин Д. А.* Геометрический метод построения сбалансированных k-значных пороговых функций и синтез подстановок на их основе // Образовательные ресурсы и технологии. 2014. № 2(5). С. 76–80.
5. *Алферов А. П., Зубов А. Ю., Кузьмин А. С., Черемушкин А. В.* Основы криптографии. М.: Гелиос АРВ, 2001.
6. *Дертоузос М.* Пороговая логика. М.: Мир, 1967.
7. *Ефимов Н. В., Розендорн Э. Р.* Линейная алгебра и многомерная геометрия. М.: Наука, 1970.
8. *Глухов М. М., Шишков А. Б.* Математическая логика. Дискретные функции. Теория алгоритмов. М.: Лань, 2012.

REFERENCES

1. *Nikonov V. G. and Sarantsev A. V.* Metody kompaktnoy realizatsii biektivnykh otobrazheniy, zadannykh regulyarnymi sistemami odnotipnykh bulevykh funktsiy [Methods of compact realization of bijective mappings proposed by regular systems of one-type Boolean functions]. Vestnik RUDN. Ser. Prikladnaya i Komp'yuternaya Matematika, 2003, vol. 2, no. 1, pp. 94–105. (in Russian)
2. *Nikonov V. G. and Sarantsev A. V.* Postroenie i klassifikatsiya regulyarnykh sistem odnotipnykh funktsiy [The construction and classification of the regular systems of one-type functions]. Proc. XXXI Int. conf. "Informatsionnye tekhnologii v nauke, obrazovanii, telekommunikatsii i biznese". Moscow, 2004, vol. 5, pp. 173–174. (in Russian)
3. *Nikonov V. G. and Sidorov E. S.* O sposobe postroeniya vzaimno odnoznachnykh otobrazheniy pri pomoshchi kvaziadamarovykh matrits [About the possibility of one-to-one mappings representation by the quasi-hadamard matrixes]. Vestnik Moskovskogo Gosudarstvennogo Universiteta Lesa — Lesnoy Vestnik, 2009, no. 2(65), pp. 155–157. (in Russian)
4. *Nikonov V. G. and Soshin D. A.* Geometricheskiy metod postroeniya sbalansirovannykh k-znachnykh porogovykh funktsiy i sintez podstanovok na ikh osnove [The geometric method for constructing a balanced k-valued threshold functions and construction of substitutions based on them]. Obrazovatel'nye Resursy i Tekhnologii, 2014, no. 2(5), pp. 76–80. (in Russian)
5. *Alferov A. P., Zubov A. Yu., Kuz'min A. S., and Cheremushkin A. V.* Osnovy kriptografii [Basics of Cryptography]. Moscow, Gelios ARV Publ., 2001. (in Russian)
6. *Dertouzos M.* Porogovaya logika [Threshold Logic]. Moscow, Mir Publ., 1967. (in Russian)
7. *Efimov N. V. and Rozendorn E. R.* Lineynaya algebra i mnogomernaya geometriya [Linear Algebra and Multidimensional Geometry]. Moscow, Nauka Publ., 1970. (in Russian)
8. *Glukhov M. M. and Shishkov A. B.* Matematicheskaya logika. Diskretnye funktsii. Teoriya algoritmov [Mathematical Logic. Discrete Functions. Algorithms Theory]. Moscow, Lan' Publ., 2012.

МАТЕМАТИЧЕСКИЕ МЕТОДЫ КРИПТОГРАФИИ

УДК 519.23

МАТРИЧНАЯ ФОРМУЛА ДЛЯ РАСПРЕДЕЛЕНИЯ ВЫХОДА БЛОЧНОЙ СХЕМЫ ШИФРОВАНИЯ И СТАТИСТИЧЕСКИЙ КРИТЕРИЙ НА ЕЁ ОСНОВЕ

О. В. Денисов, Р. А. Былина

ООО «Центр сертификационных исследований», г. Москва, Россия

Рассматривается произвольная блочная итеративная схема шифрования со случайными независимыми двоичными входными и ключевыми векторами. С помощью псевдобулевого линейного представления итерационной вектор-функции получена матричная формула для спектра распределения выхода. На основе формулы построен статистический критерий проверки гипотезы о том, что наблюдаемые двоичные векторы получены как выход схемы, против гипотезы о равномерности их распределения; рассчитаны асимптотические вероятности ошибок. Проведено экспериментальное сравнение критерия с тестом «стопка книг» (а также с его предлагаемой модификацией) при построении атаки различения на модели блочной шифрсистемы PRESENT с длиной блока 12 битов и числом раундов $R \leqslant 10$.

Ключевые слова: *двоичная вектор-функция, блочная итеративная схема шифрования, спектр распределения, атака различения, тест «стопка книг».*

DOI 10.17223/20710410/32/3

MATRIX FORMULA FOR THE SPECTRUM OF OUTPUT DISTRIBUTION OF BLOCK CIPHER SCHEME AND STATISTICAL CRITERION BASED ON THIS FORMULA

O. V. Denisov, R. A. Bylina

Certification Research Center, Moscow, Russia

E-mail: denisovOleg@yandex.ru, bopobey@rambler.ru

Arbitrary block iterative cipher scheme with random independent binary input and output vectors is considered. A matrix formula for the spectrum of the scheme output distribution is obtained by means of the pseudo-Boolean linear representation of the iterative vector-function. Based on this formula, a statistical criterion of the hypothesis testing that binary vectors are obtained as an output of the scheme against the hypothesis of their uniform distribution is given. Asymptotic type I and type II errors are calculated. An experimental comparison of the criterion with the "Bookstack" test (and its proposed modification) is done during the construction of a distinguishing attack on the mini-models of the block cipher PRESENT (with block size 12 bits and the number of rounds $R \leqslant 10$).

Keywords: *binary vector-function, block iterative cipher scheme, spectrum of distribution, distinguishing attack, the "Bookstack" test.*

Введение

Пусть $(\mathbb{Z}_2, \oplus)$ — группа вычетов по модулю 2, $f = (f_1, \ldots, f_m) : \mathbb{Z}_2^{m+n} \to \mathbb{Z}_2^m$ — двоичная вектор-функция. Рассмотрим блочную схему, в которой из начального двоичного вектора $x = y(0) \in \mathbb{Z}_2^m$ размерности m и двоичных ключевых векторов $k(1), \ldots, k(R) \in \mathbb{Z}_2^n$ размерности n образуется последовательность

$$y(t) = f\left(y(t-1), k(t)\right) \in \mathbb{Z}_2^m, \quad 1 \leqslant t \leqslant R, \tag{1}$$

где R — количество итераций (раундов). Такие схемы могут быть частью блочной шифрсистемы, алгоритма хеширования или выработки псевдослучайной последовательности.

Будем изучать вероятностную модель, в которой

$$x, k(1), \ldots, k(R) \text{ — независимые случайные векторы} \tag{2}$$

и их распределения (в общем случае произвольные) известны. Требуется: 1) найти распределение выхода схемы $y = y(R)$; 2) в случае неравномерного распределения y построить *атаку различения* на схему (1), т. е. критерий для проверки по одинаково распределённым наблюдениям

$$\xi^{(1)}, \ldots, \xi^{(N)} \text{ — независимые случайные векторы, } \xi^{(i)} \sim \xi,\ 1 \leqslant i \leqslant N, \tag{3}$$

простой гипотезы о равномерности их распределения $H_1 : \xi \sim U(\mathbb{Z}_2^m)$ против простой гипотезы $H_2 : \xi \sim y$ о том, что они имеют распределение выходных блоков схемы.

Далее в п. 1 на основе матричного действительнозначного представления функции f получена формула: спектр распределения y представлен произведением спектра распределения x и матриц, соответствующих средним значениям матриц случайных раундовых преобразований. Она позволяет при росте R и фиксированном m вычислять распределение y с временной сложностью, линейно зависящей от R, в то время как общее число 2^{nR} наборов, составленных из раундовых ключей, растет экспоненциально.

В п. 2 с помощью формулы построен критерий проверки гипотез, имеющий заданный асимптотический размер (вероятность ошибки 1-го рода). Получены оценки объёма материала N, при котором вероятность ошибки 2-го рода асимптотически не превосходит заданного значения.

Атаки различения, построенные на базе критерия в рассматриваемых условиях, учитывают внутреннее строение шифрсистемы и распределение входных блоков. Это является большим потенциальным преимуществом критерия по сравнению с универсальными статистическими критериями (тестами). Кроме того, представляется сложной задачей теоретический расчёт вероятностей ошибок универсальных (и специальных) тестов при гипотезе о том, что наблюдаются выходные блоки шифрсистемы. Авторам известны лишь нестрогие оценки таких вероятностей, полученные при эвристических предположениях.

Для экспериментов выбраны оригинальные мини-модели [1] с длиной блока $m = 8, 12$ известной легковесной шифрсистемы PRESENT, в которых алгоритм развёртывания ключа шифрования был заменён случайным неравновероятным выбором раундовых ключей, число раундов $R \leqslant 10$. В п. 3 излагается методика и результаты проведения атак различения, построенных на основе нашего критерия, а также на основе универсального теста «стопка книг» и предложенной его модификации.

1. Матричные формулы

1.1. Линейное представление двоичного отображения

Для булевых функций $g : \mathbb{Z}_2^n \to \mathbb{Z}_2$ известно [2, с. 78] псевдобулево представление в виде суммы

$$(-1)^{g(x)} = \sum_{a \in \mathbb{Z}_2^n} C(a, g)(-1)^{\langle a,x \rangle}, \quad \langle a, x \rangle = a_1 x_1 \oplus \ldots \oplus a_n x_n,$$

где $C(a, g) = 2^{-n} \sum\limits_{x \in \mathbb{Z}_2^n} (-1)^{\langle a,x \rangle + g(x)}$ — спектральный коэффициент функции g (нормированный коэффициент Уолша — Адамара).

Вектор $C(g) = \big(C(a, g) : a \in \mathbb{Z}_2^n\big)$ размерности 2^n называется *спектром булевой функции* g. Например, спектр функции-константы 0 равен $e_1 = (1, 0, \ldots, 0)$, где левая координата соответствует нулевому вектору. Здесь и далее через $e_i = (\underbrace{0, \ldots, 0}_{i-1}, 1, 0, \ldots, 0)$ обозначаем векторы стандартного базиса длины, определяемой контекстом.

Для двоичного вектора $x \in \mathbb{Z}_2^n$ через

$$w(x) = (w_a(x) : a \in \mathbb{Z}_2^n), \quad w_a(x) = (-1)^{\langle a,x \rangle},$$

обозначим соответствующий ему расширенный действительнозначный вектор, состоящий из псевдобулевых образов линейных комбинаций компонент x. Это даёт следующую краткую запись псевдобулева представления булевой функции:

$$(-1)^{g(x)} = w(x)C(g)^{\downarrow}, \quad x \in \mathbb{Z}_2^n. \tag{4}$$

Здесь и далее через $a^{\downarrow} = (a)^{\mathrm{T}}$ обозначаем вектор-столбцы.

Теперь перейдём к произвольной двоичной вектор-функции (булевому отображению) $g = (g_1, \ldots, g_m) : \mathbb{Z}_2^n \to \mathbb{Z}_2^m$. Для неё представление (4) принимает вид

$$\text{если } y = g(x), \text{ то } w(y) = w(x)\Psi(g), \tag{5}$$

где $\Psi(g) = \|C(a, \langle b, g(x) \rangle)\|_{a \in \mathbb{Z}_2^n, b \in \mathbb{Z}_2^m} = \Big(e_1^{\downarrow}, C(g_1)^{\downarrow}, \ldots, C(g_1 \oplus \ldots \oplus g_m)^{\downarrow}\Big)$ — матрица размера $2^n \times 2^m$. Такие матрицы возникают во многих криптографических приложениях [3; 4, гл. 7; 5]. Тогда, как правило, $m = n$ и функции g являются биективными S-блоками.

При $m = n$ критерием биективности функции g является сбалансированность всех ненулевых линейных комбинаций её компонент, что эквивалентно равенству $\Psi(g)_0 = e_1$, а также ортогональности матрицы $\Psi(g)$ [3, с. 279; 5, теорема 3]. Заметим, что ранее А. С. Амбросимов доказал более общий критерий того, что вектор-функция над полем Галуа сохраняет равномерное распределение [6, теорема 3].

Для функции усложнения f схемы (1), аргумент которой состоит из двух частей, получим другое представление, которое даст возможность усреднить матрицы по случайной части k. Далее обозначим через $\|$ операцию конкатенации векторов.

Лемма 1. Если $y = f(x, k)$, то $w(y) = w(x)\Psi(f, w(k))$, где $\Psi(f, w)$ — матрица размера $2^m \times 2^m$ с элементами

$$\Psi(f, w)_{a,b} = \sum_{c \in \mathbb{Z}_2^n} C\left(a \| c, \langle b, f(x, k) \rangle\right) w_c, \quad w \in \mathbb{R}^{2^n}. \tag{6}$$

Доказательство. Обозначая $C(u,b) = C\left(u, \langle b, f(x,k)\rangle\right)$ для краткости, имеем представление

$$w_b(y) = \sum_{u \in \mathbb{Z}_2^{m+n}} C(u,b) w_u(x\|k).$$

Полагая $u = a\|c$, с учётом равенства $w_u(x\|k) = w_a(x)w_c(k)$ получаем

$$w_b(y) = \sum_{a \in \mathbb{Z}_2^m} w_a(x) \left(\sum_{c \in \mathbb{Z}_2^n} C(a\|c, b) w_c(k) \right) = w(x)\Psi(f, w(k))_b^{\downarrow},$$

что и требовалось доказать. ■

Из леммы 1 следует, что в схеме (1)

$$w(y) = w(x)\,\Psi(f, w(k(1)))\ldots\Psi(f, w(k(R))). \tag{7}$$

Поэтому $\Psi(f, w(k(r)))$ является матрицей связи входа и выхода r-го раунда.

1.2. Спектр распределения выходного случайного вектора

Преобладанием нуля в распределении двоичной случайной величины η называется число

$$\mathsf{E}(-1)^{\eta} = \mathsf{P}\{\eta = 0\} - \mathsf{P}\{\eta = 1\} = 2\,\mathsf{P}\{\eta = 0\} - 1 \in [-1, 1].$$

Спектром распределения двоичного случайного вектора ξ называется

$$\phi(\xi) = \mathsf{E}w(\xi)$$

— вектор, состоящий из преобладаний в распределениях всевозможных линейных комбинаций компонент вектора ξ.

Найдём формулу связи между спектрами входа и выхода одной итерации, а затем — формулу связи спектров входа и выхода всей схемы.

Теорема 1. Пусть в схеме (1) выполнено условие (2) независимости входа и ключевых векторов. Тогда

$$\phi\big(y(t)\big) = \phi\big(y(t-1)\big)\,\Psi\big(f, \phi(k(t))\big), \quad 1 \leqslant t \leqslant R,$$
$$\phi\big(y\big) = \phi(x)\,\Psi\big(f, \phi(k(1))\big)\ldots\Psi\big(f, \phi(k(R))\big).$$

Доказательство. Согласно лемме 1, для каждого $1 \leqslant t \leqslant R$

$$w(y(t)) = w(y(t-1))\Psi(f, w(k(t))). \tag{8}$$

Легко доказать, что если случайные матрицы $B_1, \ldots, B_l$ независимы (рассматриваем их как случайные векторы), то $\mathsf{E}(B_1 \ldots B_l) = \mathsf{E}B_1 \ldots \mathsf{E}B_l$. Как и ранее, под случайными вектором и матрицей мы понимаем наборы случайных величин, заданных на одном вероятностном пространстве.

Случайный вектор $y(t-1)$ не зависит от случайной матрицы $\Psi(f, w(k(t)))$, поскольку они являются соответственно функциями от независимых наборов случайных векторов $x, k(1), \ldots, k(t-1)$ и $\{k(t)\}$. Поэтому первая формула получается путём взятия математического ожидания от обеих частей (8) с учётом того, что $\mathsf{E}\Psi(f, w) = \Psi(f, \mathsf{E}w)$ для любого случайного действительнозначного вектора w.

Вторая формула аналогично вытекает из равенства (7). ■

Таким образом, матрица $\Psi(f, \phi(k(r)))$, полученная интегрированием (взятием математического ожидания) случайной матрицы $\Psi(f, w(k(r)))$, является *матрицей связи спектров* входа и выхода на r-м раунде.

Отметим, что спектр $\phi = \phi(\xi)$ распределения является набором всех значений характеристической функции [6, 7] случайного двоичного вектора $\xi = (\xi_1, \ldots, \xi_m)$ и поэтому полностью определяет распределение ξ [7, с. 102]. В более простом виде для случая группы вычетов $\mathbb{Z}_k$ формула обращения приведена в [8, с. 25]. Через спектр выражается, в частности, ковариационная матрица расширенного случайного вектора $w(\xi)$:

$$\Sigma\big(w(\xi)\big) = \|\phi_{a\oplus b} - \phi_a\phi_b\|_{a,b\in\mathbb{Z}_2^m}. \tag{9}$$

Эта формула легко следует из равенства $w_a(\xi)w_b(\xi) = w_{a\oplus b}(\xi)$.

1.3. Матрицы связи в случае покоординатного наложения раундовых ключей

Рассмотрим распространённую (в частности, используемую в PRESENT) схему аддитивного наложения раундовых ключей. Здесь $n = m$ и ключевой вектор накладывается посредством векторного сложения, то есть

$$f(x, k) = g(x \oplus k), \quad g : \mathbb{Z}_2^m \to \mathbb{Z}_2^m. \tag{10}$$

В этом случае выражение (6) для матрицы $\Psi(f, w)$ упрощается.

Теорема 2. Если функция f имеет вид (10), то столбцы матрицы связи равны

$$\Psi(f, w)_b^{\downarrow} = C(\langle b, g(x)\rangle)^{\downarrow} * w^{\downarrow}, \quad b \in \mathbb{Z}_2^m, \tag{11}$$

где $*$ — операция покоординатного умножения векторов.

Доказательство. Зафиксируем вектор b и, делая замену $y = x \oplus k$ в выражении для спектральных коэффициентов, имеем

$$C(a\|c, \langle b, f\rangle) = 2^{-2m} \sum_{x\in\mathbb{Z}_2^m, k\in\mathbb{Z}_2^m} (-1)^{\langle a,x\rangle+\langle c,k\rangle+\langle b,g(x\oplus k)\rangle} =$$

$$= 2^{-m} \sum_{x\in\mathbb{Z}_2^m} (-1)^{\langle a,x\rangle} \cdot 2^{-m} \sum_{y\in\mathbb{Z}_2^m} (-1)^{\langle c,x\oplus y\rangle+\langle b,g(y)\rangle} =$$

$$= 2^{-m} \sum_{x\in\mathbb{Z}_2^m} (-1)^{\langle a\oplus c,x\rangle} C(c, \langle b, g\rangle) = \begin{cases} 0, & a \neq c, \\ C(c, \langle b, g\rangle), & a = c. \end{cases}$$

Таким образом, из 2^{2m} коэффициентов $C(a\|c, \langle b, f(x, k)\rangle)$ ненулевыми могут быть лишь 2^m коэффициентов, у которых $a = c$. Тогда из леммы 1 получаем, что $\Psi(f, w)_{a,b} = C(a, \langle b, g\rangle)w_a$. ■

2. Критерий проверки гипотез

2.1. Спектральная формулировка гипотез

Справедлив спектральный критерий равномерности распределения случайного вектора [7, с. 102] (см. также более простую форму [8, с. 26]), который для двоичного вектора принимает вид

$$\xi \sim U(\mathbb{Z}_2^m) \iff \phi_a(\xi) = \mathbb{I}\{a = 0\}, \ a \in \mathbb{Z}_2^m, \tag{12}$$

где $\mathbb{I}\{A\}$ — индикатор выполнения условия A. Поэтому гипотеза H_1 эквивалентна тому, что $\phi(\xi) = e_1$. При этом

$$\Sigma\big(w(\xi)\big) = \operatorname{diag}(0, 1, \ldots, 1), \tag{13}$$

так как, согласно (9) и (12), $\Sigma_{a,b} = \mathbb{I}\{a = b\} - \mathbb{I}\{a = 0\}\mathbb{I}\{b = 0\}$. Левый верхний элемент ковариационной матрицы равен 0, поскольку $w_0(\xi) \equiv 1$.

Далее считаем, что распределение выхода схемы неравномерно, т. е.

$$\phi(y) \neq e_1, \tag{14}$$

что даёт принципиальную возможность построения атаки различения. Получим условие, необходимое для (14).

Лемма 2. Если $k \sim U(\mathbb{Z}_2^n)$, то f биективна тогда и только тогда, когда $\Psi(f, \phi(k))_0 = e_1$.

Доказательство. Из леммы 1 с учётом критерия (12) и вышеупомянутого критерия [3] биективности функции получаем, что элементы верхней строки матрицы $\Psi = \Psi(f, w(k))$ равны

$$\Psi_{0,b} = \sum_{c \in \mathbb{Z}_2^n} C\left(0 \| c, \langle b, f(x,k)\rangle\right) \mathbb{I}\{c = 0\} = C\left(0, \langle b, f(x,k)\rangle\right) = \mathbb{I}\{b = 0\},$$

что и требовалось доказать. ■

Из теоремы 1 и леммы 2 следует, что в случае биективной функции f (что часто выполнено для функций усложнения блочных шифрсистем) для (14) необходимо, чтобы

$$\phi(x) \neq e_1 \text{ или } \phi(k(t)) \neq e_1 \text{ для некоторого } 1 \leqslant t \leqslant R, \tag{15}$$

т. е. чтобы вход или хотя бы один ключевой вектор были распределены неравномерно.

2.2. Построение и расчёт критерия

Пусть далее E — единичная матрица, размер которой определяется контекстом, $\Phi(x)$ — функция распределения стандартного нормального закона $\mathcal{N}(0,1)$, κ_γ — квантиль уровня γ распределения $\mathcal{N}(0,1)$, т. е. $\Phi(\kappa_\gamma) = \gamma$. Нам потребуется следующее вспомогательное утверждение о многомерных нормальных распределениях.

Лемма 3. Пусть $\alpha \in (0,1)$, $\mu \neq 0$, $\nu = \mu/|\mu|$ — вектор евклидовой нормы 1, сонаправленный с μ. Тогда

$$\mathsf{P}\{\eta\,\nu^{\downarrow} \leqslant \kappa_{1-\alpha}\} = \begin{cases} 1-\alpha \text{ при } \eta \sim \mathcal{N}(0, E), \\ \Phi\left(\dfrac{\kappa_{1-\alpha} - |\mu|}{\sigma}\right) \text{ при } \eta \sim \mathcal{N}(\mu, \Sigma) \end{cases}$$

для любой ковариационной матрицы Σ, вырожденной или невырожденной, если $\sigma^2 = \nu\,\Sigma\nu^{\downarrow} > 0$.

Доказательство. Известно, что если η имеет нормальное распределение, вырожденное или невырожденное, то любая линейная комбинация его компонент $\zeta = \eta\,\nu^{\downarrow}$ имеет нормальное распределение со средним $(\mathsf{E}\eta)\,\nu^{\downarrow}$ и дисперсией $\nu\Sigma\nu^{\downarrow}$. Тогда при $\eta \sim \mathcal{N}(0, E)$ имеем $\mathsf{E}\zeta = 0$, $\mathsf{D}\zeta = \nu\nu^{\downarrow} = |\nu|^2 = 1$, и искомая вероятность равна $\Phi(\kappa_{1-\alpha}) = 1 - \alpha$. При $\eta \sim \mathcal{N}(\mu, \Sigma)$ имеем $\mathsf{E}\zeta = \mu\nu^{\downarrow} = |\mu|$, $\mathsf{D}\zeta = \nu\,\Sigma\nu^{\downarrow} = \sigma^2$, и вероятность равна $\mathsf{P}\left\{(\zeta - |\mu|)/\sigma \leqslant (\kappa_{1-\alpha} - |\mu|)/\sigma\right\} = \Phi\left((\kappa_{1-\alpha} - |\mu|)/\sigma\right)$. ■

Для двоичного вектора $x \in \mathbb{Z}_2^m$ через $\tilde{w}(x) = (w_a(x) : 0 \neq a \in \mathbb{Z}_2^m)$ обозначим действительнозначный вектор размерности $2^m - 1$, полученный удалением из $w(x)$ крайней левой компоненты, тождественно равной 1, и рассмотрим статистику

$$S(N) = \tilde{w}\left(\xi^{(1)}\right) + \ldots + \tilde{w}\left(\xi^{(N)}\right)$$

— сумму независимых одинаково распределённых случайных векторов согласно условию (3).

Согласно центральной предельной теореме [9, с. 435], при $N \to \infty$ распределение нормированной статистики сходится к нормальному:

$$\frac{1}{\sqrt{N}}(S(N) - N\mathsf{E}\tilde{w}(\xi)) \xrightarrow{D} \mathcal{N}(0, \Sigma(\tilde{w}(\xi))).$$

Отсюда с учётом (12) и (13) следует, что при гипотезе H_1 предельное распределение будет стандартным нормальным: $\dfrac{S(N)}{\sqrt{N}} \xrightarrow{D} \mathcal{N}(0, E)$.

Итак, получаем критерий

$$S(N)\,\nu^{\downarrow} > \sqrt{N}\kappa_{1-\alpha} \Longrightarrow H_2, \tag{16}$$

где $\nu = \tilde{\phi}(y)/|\tilde{\phi}(y)|$; $\tilde{\phi}(y)$ — вектор, полученный из $\phi(y)$ отбрасыванием левой координаты. Его критической областью является полупространство, отделённое от начала координат гиперплоскостью, перпендикулярной вектору, направленному от 0 к $\tilde{\phi}(y)$.

Из леммы 3 и теоремы непрерывности 2 [9, с. 32] следует, что критерий имеет асимптотический размер α, т. е. $\alpha_1(N) \to \alpha$ при $N \to \infty$ для любого фиксированного $\alpha \in (0, 1)$.

Оценим вероятность ошибки второго рода. При гипотезе H_2 и больших N распределение $\dfrac{S(N)}{\sqrt{N}}$ близко к $\mathcal{N}(\sqrt{N}\tilde{\phi}(y), \tilde{\Sigma})$, где $\tilde{\Sigma} = \Sigma(\tilde{\phi}(y)))$ может быть получена из матрицы (9), вычисленной при $\xi = y$, путём отбрасывания верхней строки и левого столбца. Тогда по лемме 3

$$\alpha_2(N) \approx \Phi\left(\frac{\kappa_{1-\alpha} - \sqrt{N}|\tilde{\phi}(y)|}{\sigma}\right), \text{ где } \sigma^2 = \nu\,\tilde{\Sigma}\nu^{\downarrow}.$$

Следовательно, значение $\alpha_2(N)$ будет близко к заданному значению $\beta \in (0, 1)$ при

$$N = N^*(\alpha, \beta) = \frac{(\kappa_{1-\alpha} + \kappa_{1-\beta}\sigma)^2}{|\tilde{\phi}(y)|^2}. \tag{17}$$

3. Атака различения на SP-сеть с независимыми раундовыми ключами

Критерий для проверки гипотезы H_1 против H_2 в криптографической литературе называется *атакой различения* на заданную схему. Для построения критерия требуется знание распределения y выхода схемы. Рассмотрим возможные способы его вычисления. Если ключи $k(t)$ извлекаются равновероятно из некоторых ключевых множеств $K(t) \subset \mathbb{Z}_2^n$, а x из $X \subset \mathbb{Z}_2^m$, то поиск распределения выхода путём тотального вычисления всех значений $y = y(x, k(1), \ldots, k(R))$ имеет сложность, пропорциональную $|X| \cdot |K(1)| \cdot \ldots \cdot |K(R)|$. Если все ключевые множества имеют мощность K, то с ростом R эта сложность растет как K^R, т. е. экспоненциально. Другой путь поиска распределения даёт матричная формула теоремы 1: здесь временная сложность равна сложности вычисления произведения R матриц, что при росте R и фиксированном m оценивается величиной, линейно зависящей от R. Справедливости ради заметим, что при этом способе сравнительно большое значение имеет емкостная сложность хранения матриц Ψ, пропорциональная величине 2^{2m}.

Заметим также, что существует ещё путь статистического оценивания распределения y при случайном выборе $x, k(1), \ldots, k(R)$, но его исследование выходит за рамки данной работы.

Перейдём к выбору схемы шифрования, а также распределений входного блока и раундовых ключей. Схемы шифрования были выбраны из семейства SmallPresent[m] [1] масштабируемых моделей шифрсистемы PRESENT, у которых длина блока m, измеряемая в битах, кратна 4. Так как шифрсистема является SP-сетью с покоординатным наложением раундовых ключей, то $m = n$ и матрица связи вычисляется по формуле (11). Из-за большого размера $2^m \times 2^m$ матрицы Ψ мы были вынуждены ограничиться сравнительно небольшими значениями $m \leqslant 12$. Оригинальный алгоритм получения раундовых ключей заменён на описываемую ниже схему (19) их случайной генерации, что назовём *неавтономной моделью*. Поскольку должно быть обеспечено условие (3) независимости наблюдаемых блоков, наборы раундовых ключей будут выбираться независимо для каждого входного блока, что соответствует ситуации *«много одноблочных сообщений»*.

Ограничимся случаем, когда все раундовые ключи распределены одинаково: $k(t) \sim k$, и тогда при условии (2)

$$\phi(y) = \phi(x)\Psi(f, \phi(k))^R. \tag{18}$$

Согласно (15), распределение x или k должно быть неравномерным. Будем использовать модель, в которой все биты этих векторов выбираются независимо с преобладанием $d \neq 0$, т. е. каждый бит имеет распределение Бернулли $\mathrm{Be}\left(\dfrac{1-d}{2}\right)$, и обозначать это так:

$$x \sim k \sim \mathrm{Be}^m\left(\frac{1-d}{2}\right). \tag{19}$$

Заметим, что тогда

$$\phi(x) = \phi(k) = (d^{\|a\|} : a \in \mathbb{Z}_2^m) = (1, d, \ldots, d^m).$$

Далее описаны эксперименты двух типов: 1) построение критериев (16) и оценки их характеристик; 2) сравнение их с критерием «стопка книг», далее называемым «СК-тест» для краткости.

3.1. Эксперименты по построению и применению критерия

Здесь целью экспериментов является построение критерия, проверка соответствия реально наблюдаемых ошибок критерия (16) и заданных значений α, β. Для этого проводилось $M = 50$ серий получения выборок при каждой гипотезе. Опишем методику проведения экспериментов.

1. Для заданных распределений x и k и функции f находим спектры $\phi(x), \phi(k)$, матрицу $\Psi(f, \phi(k))$, а затем спектр $\phi(y)$ по формуле (18).

2. Вычисляем $\tilde{\phi}(y)$, ν, $\tilde{\Sigma}$ и $N = N^*(\alpha, \beta)$ для заданных значений α, β согласно (17).

3. Генерируем независимо M выборок $\xi^{(t)}$, $1 \leqslant t \leqslant M$, объёма N по случайной равновероятной схеме, применяем к каждой выборке критерий. Получаем значение $\hat{\alpha} = M_2/M$, где M_2 — количество решений в пользу H_2.

4. Генерируем независимо M серий по N независимых наборов $(x, k(1), \ldots, k(R))$ в соответствии с заданными распределениями, применяем к каждому набору схему шифрования, получаем выборку $\xi^{(t)}$ из N выходных векторов схемы. Применяя к каждой выборке критерий, находим значение $\hat{\beta} = M_1/M$, где M_1 — количество решений в пользу H_1.

Здесь взято число раундов $R = 10$ и значения $\alpha = \beta = 0{,}2$ выбраны достаточно большими для того, чтобы при небольшом числе $M = 50$ серий математическое ожидание значения $\hat{\alpha}$ было не очень мало. Графики зависимости $N^*(\alpha, \beta)$ от значений преобладания d представлены на рис. 1.

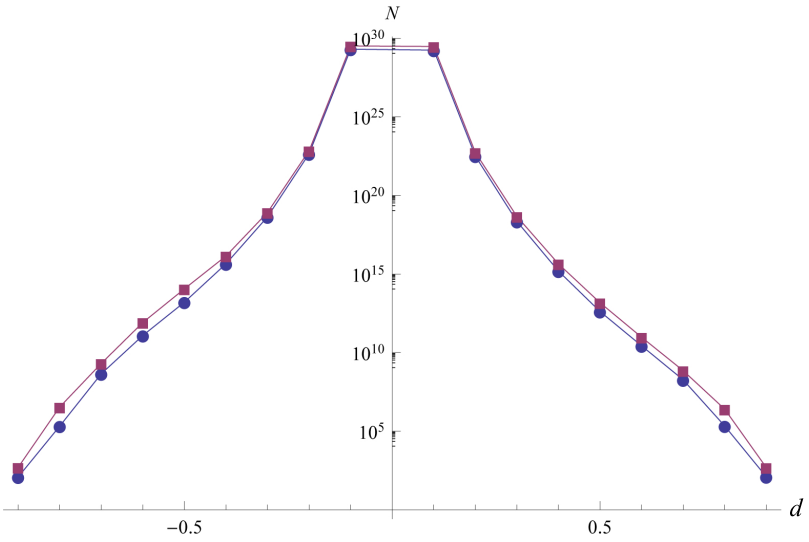

Рис. 1. Зависимость $N^*(\alpha, \beta)$ от d для моделей SmallPresent$[m]$, $m = 8, 12$

Заметим, что на рис. 1 вычисленные значения $N^*(\alpha, \beta)$ монотонно зависят от $|\tilde{\phi}(y)|$. Это, вообще говоря, не следует из (17), поскольку в числителе этой формулы σ также зависит от $\tilde{\phi}(y)$.

Численные результаты экспериментов для некоторых значений преобладания d представлены в табл. 1. Реальные вероятности шестнадцати полученных ошибок критерия (16) лежат в пределах от 0,06 до 0,3, что согласуется с теорией.

Таблица 1

Результаты экспериментов с неавтономными моделями SmallPresent$[m]$ при $R = 10$, $M = 50$, $\alpha = \beta = 0{,}2$, $x \sim k \sim \mathrm{Be}^m\left(\dfrac{1-d}{2}\right)$ в ситуации одноблочных сообщений

Параметры	SmallPresent[8]				SmallPresent[12]			
d	-0,9	-0,8	0,8	0,9	-0,9	-0,8	0,8	0,9
$\lvert\tilde{\phi}(y)\rvert$	0,21	$3{,}6 \cdot 10^{-3}$	$3{,}63 \cdot 10^{-3}$	0,20	0,11	$9{,}08 \cdot 10^{-4}$	$1{,}0 \cdot 10^{-3}$	0,11
σ^2	3,01	1,69	0,99	3,02	3,73	1,0	1,0	3,66
$N_1(\alpha, \beta, d)$	121	$2{,}7 \cdot 10^5$	$2{,}14 \cdot 10^5$	127	501	$3{,}4 \cdot 10^6$	$2{,}5 \cdot 10^6$	492
$\widehat{\alpha}$	0,22	0,06	0,22	0,24	0,24	0,30	0,10	0,08
$\widehat{\beta}$	0,20	0,20	0,16	0,12	0,16	0,06	0,10	0,18

Так как при уменьшении $|d|$ распределение $\mathrm{Be}^m\left(\dfrac{1-d}{2}\right)$ приближается к равномерному, т. е. гипотезы сближаются, то значение $|\tilde{\phi}(y)|$ уменьшается, а объём материала $N^*(\alpha, \beta, d)$ увеличивается. Поэтому для $|d| \leqslant 0{,}7$ статистические эксперименты не проводились.

3.2. Эксперименты для сравнения критерия и СК-теста

СК-тест и предлагаемая его модификация

На основе адаптивной структуры «стопка книг», которая предложена Б. Я. Рябко в 1980 г., позже был построен СК-тест. Он является одним из мощных универсальных тестов для проверки гипотезы о равновероятности исходов в статистических моделях с S исходами, где S значительно больше объёма выборки. С 2004 г. он применялся в ряде работ [10 – 13] для оценки качества псевдослучайных последовательностей, вырабатываемых поточными шифрсистемами, а также блочными шифрсистемами в режиме CTR — так называется режим работы, в котором входной блок с номером $t = 0, 1, \ldots$ является m-битовым двоичным представлением числа t, т. е. это режим «счётчика».

Далее ограничимся часто применяемой версией СК-теста, в которой множество X исходов, $|X| = S$, разбивается на два множества, и первое из них (верхняя часть стопки) имеет мощность $Q \ll S$. В наших обозначениях эта версия описывается так: фиксируется A_1 — произвольный список, содержащий Q различных элементов X. Далее для наблюдаемой случайной последовательности блоков $\xi^{(1)}, \ldots, \xi^{(N)}$ рекуррентно строится последовательность $\{A_t\}$ случайных списков мощности Q, где A_{t+1} определяется по A_t и $\xi^{(t)}$ следующим образом. Элемент $\xi^{(t)}$ становится первым элементом A_{t+1} и к нему присоединяется список A'_t мощности $Q-1$, где A'_t получен из A_t удалением $\xi^{(t)}$ при $\xi^{(t)} \in A_t$ либо удалением последнего элемента A_t при $\xi^{(t)} \notin A_t$. Рассматривается статистика

$$\nu = \sum_{1 \leqslant t \leqslant N} \mathbb{I}\{\xi^{(t)} \in A_t\}.$$

Лемма 4. При гипотезе H_1 о независимости и равновероятности блоков события $\{\xi^{(t)} \in A_t\}$ независимы и статистика имеет биномиальное распределение $\nu \sim \mathrm{Bin}(N, q)$, $q = Q/S$.

Доказательство. Заметим, что при фиксированном значении A_1 фиксация значений $\xi^{(1)} = a_1, \ldots, \xi^{(t-1)} = a_{t-1}$ однозначно определяет значение случайного списка A_t, $t \geqslant 1$, которое обозначим через

$$B_t = B_t(A_1, a_1, \ldots, a_{t-1}).$$

Тогда для любых $k \geqslant 1$ фиксированных номеров $1 \leqslant t_1 < \ldots < t_k = T$, обозначая $l = T - k$, $\{s_1, \ldots, s_l\} = \{1, \ldots, T\} \setminus \{t_1, \ldots, t_k\}$, имеем

$$\mathsf{P}\left\{\xi^{(t_1)} \in A_{t_1}, \ldots, \xi^{(t_k)} \in A_{t_k}\right\} =$$

$$= \sum_{a_{s_1}, \ldots, a_{s_l} \in X} \mathsf{P}\left\{\xi^{(t_1)} \in A_{t_1}, \ldots, \xi^{(t_k)} \in A_{t_k}, \xi^{(s_1)} = a_{s_1}, \ldots, \xi^{(s_l)} = a_{s_l}\right\} =$$

$$= \sum_{a_{s_1}, \ldots, a_{s_l} \in X} \sum_{a_{t_1} \in B_{t_1}} \cdots \sum_{a_{t_k} \in B_{t_k}} \mathsf{P}\left\{\xi^{(1)} = a_1, \ldots, \xi^{(T)} = a_T\right\} = S^l Q^k \frac{1}{S^T} = \left(\frac{Q}{S}\right)^T = q^k,$$

поскольку все вероятности в последней сумме равны $1/S^T$ и каждое множество B_{t_i} содержит ровно Q элементов.

При $k = 1$ получаем $\mathsf{P}\left\{\xi^{(t)} \in A_t\right\} = q$ для всех $t \geqslant 1$, и поэтому

$$\mathsf{P}\left\{\xi^{(t_1)} \in A_{t_1}, \ldots, \xi^{(t_k)} \in A_{t_k}\right\} = \mathsf{P}\left\{\xi^{(t_1)} \in A_{t_1}\right\} \ldots \mathsf{P}\left\{\xi^{(t_k)} \in A_{t_k}\right\}.$$

Лемма 4 доказана. ■

В упомянутых работах рассматривается статистика типа «хи-квадрат»

$$x^2 = \frac{(\nu - Nq)^2}{Nq} + \frac{((N-\nu) - N(1-q))^2}{N(1-q)} = \frac{(\nu - Nq)^2}{Nq(1-q)} = \tilde{\nu}^2. \tag{20}$$

Она является квадратом случайной величины $\tilde{\nu} = \dfrac{\nu - Nq}{\sqrt{Nq(1-q)}}$, распределение которой с учётом леммы 4 сходится к $\mathcal{N}(0,1)$ при $N \to \infty$ и фиксированном q. Следовательно, $x^2 \xrightarrow{D} \chi_1^2$ при гипотезе H_1, что неоднократно использовалось, но строгого доказательства этого факта нам найти не удалось.

В [14] изучается, а в [12, 13] применяется следующий критерий согласия с H_1: если $x^2 \geqslant \kappa(1-\alpha, \chi_1^2)$, где κ — квантиль распределения χ_1^2, то H_1 отклоняется. Он имеет асимптотический размер α и эквивалентен двустороннему критерию

$$\text{если } |\tilde{\nu}| \geqslant \kappa_{1-\alpha/2}, \text{ то } H_1 \text{ отклоняется}, \tag{21}$$

где κ_γ — квантиль распределения $\mathcal{N}(0,1)$, как и раньше.

Как отмечается в [10, с. 74], при альтернативе о неравновероятности исходов в выборке чаще появляются более вероятные исходы, и они проводят в верхней части стопки значительно больше времени, чем остальные. Но это всегда ведет к смещению распределения ν вправо от Nq, и, по нашему мнению, при независимых наблюдениях более адекватен следующий односторонний критерий:

$$\text{если } \nu \geqslant Nq + \kappa_{1-\alpha}\sqrt{Nq(1-q)}, \text{ то } H_1 \text{ отклоняется} \tag{22}$$

асимптотического размера α. Этот модифицированный СК-тест для краткости далее называем *МСК-тестом*.

Заметим, что двусторонний критерий (21) также выявляет смещение распределения ν влево от Nq. Такое смещение возникает, например, когда наблюдения зависимы и появление элемента в выборке уменьшает эмпирическую вероятность его появления. В частности, это соответствует альтернативе, при которой выбор элементов происходит случайно без возвращения.

При гипотезе типа H_2 строгий расчёт тестов является сложной задачей, поскольку в общем случае распределение статистик будет зависеть от конкретного ключа (например, в экспериментах [11, табл. 1] эмпирическая ошибка 2-го рода адаптивного теста χ^2 при его применении к выходным последовательностям длины $N = 2^{20}$ блоков трехраундового RC5 ($m = 64$) изменялась от 0,95 до 0,01 в зависимости от ключа шифрования; эксперименты проводились для 10 случайных ключей). Для универсальных тестов, вероятно, задача ещё более усложняется, поскольку они не «подстроены» под конкретную шифрсистему; авторы не встречались с примерами её строгого решения. Это главная причина отсутствия теоретического сравнения вероятностей ошибок 2-го рода нашего критерия с другими тестами, и далее проводится его экспериментальное сравнение с СК- и МСК-тестами.

Здесь можно добавить, что в [14] изучается распределение статистики x^2 при альтернативе $H^{\gamma,\delta}$, $\gamma, \delta \in (0,1)$, заключающейся в том, что в схеме с S исходами некоторые $S\gamma$ исходов имеют вероятность $(1+\delta)/S$, некоторые $S\gamma$ исходов имеют вероятность $(1-\delta)/S$, а оставшиеся $S(1-2\gamma)$ исходов имеют вероятность $1/S$. Получена следующая асимптотическая оценка для объёма выборки: для любых $\alpha, \beta \in (0,1)$ существует

такое $C > 0$, что при $S \to \infty$ вероятности ошибок СК-теста (21) с $|A_1| = [\sqrt{S}]$ асимптотически не превосходят α и β соответственно, если

$$N = C[\sqrt{S}]. \tag{23}$$

Признавая силу и универсальность СК-теста, считаем необходимым сделать следующие замечания по поводу методики его применения [13].

1. Если главная цель [13] — изучение проблемы создания быстрого генератора псевдослучайных последовательностей на базе блочных шифрсистем («сокращение числа раундов увеличит производительность шифров и позволит генерировать псевдослучайные числа быстрее» [13, с. 66]), то представляется не вполне обоснованным: а) выбор слабой CTR-последовательности в качестве входа (можно рассмотреть, например, более сильный и простой режим, при котором входным блоком шифрсистемы является предыдущий выходной блок); б) ограничение только первым выходным 32-битным словом (из 2–4 возможных в зависимости от $64 \leqslant m \leqslant 128$).

2. Как отмечалось выше со ссылкой на [10, с. 74], неравновероятность исходов в выборке смещает распределение ν вправо от Nq (применяемая версия СК-теста сводится к этой статистике согласно равенствам (20)). Но режим CTR в силу биективности блочного преобразования эквивалентен выбору без возвращения, что должно было привести к значительному (при любом числе раундов!) смещению распределения ν влево. Редуцирование выходного блока до одного слова размывает этот эффект (это, возможно, является объяснением обстоятельства $1, б$), и неясно, на что сильнее в итоге реагирует СК-тест, т. е. куда сместится распределение ν. Ответ на этот вопрос могли бы дать средние экспериментальные значения $\tilde{\nu}$. Если они большие (по абсолютной величине) отрицательные, то построенный в [13] критерий реагирует на выбор блоков без возвращения, а не на их неравновероятность.

Эксперименты с моделью SmallPresent[12]

Проведём сравнительный экспериментальный анализ нашего критерия и СК-теста (21), МСК-теста (22) при выборе значения их параметра

$$Q = 2^{m/2} = 2^6 = 64$$

согласно условию (23), принятому также в [13]. Число раундов R растёт от 3 до 10, количество серий $M = 10$, расчётные вероятности ошибок $\alpha = 0{,}05$, $\beta = 0{,}01$, набор из R раундовых ключей случайно выбирается для каждого N_2-блочного сообщения. Как и раньше, каждый из R ключей выбирается по схеме $k \sim \mathrm{Be}^{12}\left(\dfrac{1-d}{2}\right)$, $d = 0{,}8$. В частности, при $N_2 = 1$ получаем ранее исследованную ситуацию одноблочных сообщений.

Количество сообщений $N_1 = [N^*(R, \alpha, \beta, d)/N_2]$ взято таким, чтобы общий их объём, измеряемый в блоках, был близок к расчётному объёму выборки (17) нашего критерия (табл. 2).

Входные блоки выбираются независимо и равновероятно из множества слов с восемью нулевыми старшими битами:

$$x(t) \sim U(V), \quad V = \left\{x \in \mathbb{Z}_2^{12} : \; x_1 = \ldots = x_8 = 0\right\}, \tag{24}$$

что близко к режиму CTR, но, в отличие от него, $x(1), \ldots, x(N)$ независимы и неизвестны. Этот режим выбора можно назвать *известным неравномерным распределением неизвестных входных блоков*. Значение параметра 8 в (24) выбрано таким, чтобы

Таблица 2

Объём материала, достаточный для корректной работы спектрального критерия при различном значении количества раундов R шифрсистемы SmallPresent[12]

R	3	4	5	6	7	8	9	10
$N^*(R)$	18	65	363	$2{,}7 \cdot 10^3$	$1{,}9 \cdot 10^4$	$1{,}2 \cdot 10^5$	$8{,}8 \cdot 10^5$	$7{,}3 \cdot 10^6$

максимальная доля ненулевых битов в блоке, равная $\dfrac{12-8}{12} = \dfrac{1}{3}$, была близка к максимальной доле $\dfrac{24}{64} = \dfrac{3}{8}$ ненулевых битов в блоке в экспериментах [13].

Итак, в каждой серии выполняем следующие шаги:

1) Для фиксированного $R = 3, 4, \ldots$ выбираем из табл. 2 значение $N = N^*(R, \alpha, \beta, d) = N^*(R)$. Случайно равновероятно из V выбираем N_2 блоков открытого текста и формируем из них сообщения длины N_2. Генерируем таким образом $\lceil N/N_2 \rceil$ сообщений.
2) Для каждого сообщения, согласно указанному выше распределению, выбираем случайно набор из R раундовых ключей и шифруем на этом наборе все блоки сообщения. Объединяем вместе блоки всех сообщений в единую выборку.
3) Применяем к полученной выборке спектральный критерий, СК-тест и МСК-тест.

По результатам M серий вычисляем вектор эмпирических вероятностей ошибок $\beta = (\widehat{\beta}_1, \widehat{\beta}_2, \widehat{\beta}_3)$, которые равны отношению к M количества серий, в которых соответствующий тест неправильно принял решение, а также $\widetilde{\nu}_{\text{ср}}$ — эмпирическое среднее значение статистики $\tilde{\nu}$ по сериям. Эти значения представлены в табл. 3 — вектор $\beta \cdot M$ в верхней строке и $\widetilde{\nu}_{\text{ср}}$ в нижней для каждого N_2.

Таблица 3

Результаты экспериментов в модели «много коротких сообщений» для шифрсистемы SmallPresent[12]

N_2	R							
	3	4	5	6	7	8	9	10
1	0, 9, 9	0, 8, 8	1, 9, 8	0, 9, 9	1, 10, 9	1, 9, 9	1, 10, 10	1, 9, 9
	0,20	1,98	−0,28	−0,20	0,29	−0,30	−0,14	−0,20
2	1, 8, 8	1, 5, 5	1, 2, 0	1, 0, 0	1, 0, 0	1, 0, 0	2, 0, 0	1, 0, 0
	1,50	3,98	4,49	11,4	34,4	87,57	231,9	668,1
3	3, 2, 2	1, 2, 2	0, 0, 0	1, 0, 0	1, 0, 0	0, 0, 0	1, 0, 0	0, 0, 0
	3,45	5,08	10,25	25,13	67,10	172,32	455,77	1311,49
4	1, 4, 4	2, 0, 0	1, 0, 0	0, 0, 0	1, 0, 0	1, 0, 0	1, 0, 0	0, 0, 0
	2,69	5,88	14,73	37,74	98,48	252,65	669,64	1929,11
5	1, 2, 2	4, 0, 0	1, 0, 0	0, 0, 0	0, 0, 0	1, 0, 0	0, 0, 0	0, 0, 0
	5,16	7,18	18,9	48,27	129,12	330,7	874,07	2516,76
6	4, 2, 2	2, 0, 0	1, 0, 0	1, 0, 0	1, 0, 0	0, 0, 0	0, 0, 0	2, 0, 0
	4,78	9,08	21,26	59,86	159,16	406,80	1070,59	3085,88
7	1, 2, 2	2, 0, 0	2, 0, 0	0, 0, 0	1, 0, 0	1, 0, 0	0, 0, 0	1, 0, 0
	6,30	9,88	26,03	70,1	186,34	476,00	1258,59	3628,72
8	4, 0, 0	3, 0, 0	3, 0, 0	1, 0, 0	3, 0, 0	0, 0, 0	2, 0, 0	1, 0, 0
	7,63	11,29	27,81	78,40	213,51	546,40	1440,21	4148,2
9	3, 0, 0	2, 0, 0	3, 0, 0	2, 0, 0	1, 0, 0	0, 0, 0	0, 0, 0	0, 0, 0
	8,20	13,08	32,64	88,74	240,14	611,14	1613,08	4649,34

Из табл. 3 можно сделать следующие выводы:

1) При $N_2 = 1$ для всех R, а также при $N_2 = 2$ для небольшого объёма материала (случаи $R = 3, 4$) спектральный критерий работает явно лучше обоих тестов. В других случаях СК-тест и МСК-тест работают лучше нашего критерия, а распределение статистики $\widetilde{\nu}$ смещается вправо от нуля, и значение $\widetilde{\nu}_{\text{ср}}$ растёт при увеличении R. Для объяснения причин этого эффекта требуется дополнительное теоретическое исследование тестов.
2) Значение эмпирической вероятности ошибки спектрального критерия при достаточном объёме материала остаётся близким к расчётному при всех $N_2 \geqslant 2$, если число сообщений достаточно большое. Это условие примерно соответствует значениям в табл. 3 выше главной диагонали.

Заключение

В работе получена матричная формула для спектра распределения выходных блоков итеративной схемы в зависимости от распределения входных блоков и распределения независимых раундовых (итерационных) ключей. Емкостная сложность её применения оценивается величиной $\mathrm{O}(2^{m^2})$, где m бит — длина блока. Временная сложность оценивается величиной $\mathrm{O}(R\,2^{m^2})$, где R — число раундов, и может быть значительно меньше тотальной сложности вычисления распределения выходных блоков, в ряде случаев близкой к $\mathrm{O}(2^{m^2+mR})$.

Первой особенностью вероятностной модели является случайный выбор раундовых ключей, что позволяет исключить зависимость распределения выходных блоков, а также основанной на спектре распределения атаки различения от конкретного ключа шифрования. Второй особенностью является предположение о том, что зашифрование каждого входного блока производится на своём случайном наборе раундовых ключей. Вместе это даёт возможность рассчитать оценки вероятностей ошибок атаки и объём выборки без использования каких-либо эвристических предположений. Условия атаки — известные неравномерные распределения неизвестных входных блоков и раундовых ключей. Эксперименты на мини-моделях блочной шифрсистемы PRESENT с длиной блока 12 битов и числом раундов $R \leqslant 10$ показали, что согласие с теорией сохраняется при произвольном количестве N_2 блоков, шифруемых на одном наборе раундовых ключей (несмотря на то, что при этом нарушается условие (3) независимости наблюдаемых блоков), если количество сообщений также большое.

Построенная атака различения учитывает внутреннее строение шифрсистемы и распределение входных блоков, что представляется большим потенциальным преимуществом по сравнению с универсальными статистическими критериями. Произведено экспериментальное сравнение построенного критерия с тестом «стопка книг», а также с предложенной его модификацией. Оно показало, что: 1) модификация работала не хуже оригинального теста во всех экспериментах; 2) критерий работал лучше модификации при $N_2 = 1$, а также при $N_2 \geqslant 2$ на небольшом количестве сообщений.

Авторы выражают признательность рецензенту за ряд полезных замечаний, способствовавших улучшению статьи.

ЛИТЕРАТУРА

1. *Leander G.* Small Scale Variants of the Block Cipher PRESENT. Technical University of Denmark, 2010. `http://eprint.iacr.org/2010/143.pdf`.
2. *Логачев О. А., Сальников А. А., Ященко В. В.* Булевы функции в теории кодирования и криптологии. М.: МЦНМО, 2004. 470 с.

3. *Daemen J., Govaerts R., and Vandewalle J.* Correlation matrices // FSE-1995. LNCS. 1995. V. 1008. P. 275–285.
4. *Daemen J. and Rijmen V.* The design of Rijndael: AES — the Advanced Encryption Standard. Springer, 2002. 227 p.
5. *Денисов О. В.* Статистическая оценка множества существенных аргументов двоичной вектор-функции с искаженными значениями // Матем. вопр. криптографии. 2014. Т. 5. Вып. 4. С. 41–61.
6. *Амбросимов А. С.* Свойства бент-функций q-значной логики над конечными полями // Дискретная математика. 1994. Т. 6. Вып. 3. С. 50–60.
7. *Воробьев Н. Н.* Сложение независимых случайных величин на конечных абелевых группах // Матем. сборник. 1954. Т. 34(76). Вып. 1. С. 83–126.
8. *Денисов О. В.* Вероятностные свойства двоичных отображений. Учеб.-методич. пособие. М., 2008. 80 с.
9. *Боровков А. А.* Математическая статистика. М.: Наука, 1984. 472 с.
10. *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.
11. *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.
12. *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.
13. *Пестунов А. И.* Предварительная оценка минимального числа раундов легковесных шифров для обеспечения их удовлетворительных статистических свойств // Прикладная дискретная математика. Приложение. 2015. № 8. С. 66–68.
14. *Пестунов А. И.* Теоретическое исследование свойств статистического теста «стопка книг» // Вычислительные технологии. 2006. Т. 11. № 6. С. 96–103.

REFERENCES

1. *Leander G.* Small Scale Variants of the Block Cipher PRESENT. Technical University of Denmark, 2010. `http://eprint.iacr.org/2010/143.pdf`.
2. *Logachev O. A., Sal'nikov A. A., and Yashchenko V. V.* Bulevy funktsii v teorii kodirovaniya i kriptologii [Boolean Functions in Coding Theory and Cryptology]. Moscow, MCCME Publ., 2004. (in Russian)
3. *Daemen J., Govaerts R., and Vandewalle J.* Correlation matrices. FSE-1995, LNCS, 1995, vol. 1008, pp. 275–285.
4. *Daemen J. and Rijmen V.* The design of Rijndael: AES — the Advanced Encryption Standard. Springer, 2002. 227 p.
5. *Denisov O. V.* Statisticheskaya otsenka mnozhestva sushchestvennykh argumentov dvoichnoy vektor-funktsii s iskazhennymi znacheniyami [Statistical estimation of the significant arguments set of the binary vector-function with corrupted values]. Mat. Vopr. Kriptogr., 2014, vol. 5, iss. 4, pp. 41–61. (in Russian)
6. *Ambrosimov A. S.* Svoystva bent-funktsiy q-znachnoy logiki nad konechnymi polyami [Properties of bent functions of q-valued logic over finite fields]. Diskr. Mat., 1994, vol. 6, iss. 3, pp. 50–60. (in Russian)
7. *Vorob'ev N. N.* Slozhenie nezavisimykh sluchaynykh velichin na konechnykh abelevykh gruppakh [Addition of independent random variables on finite abelian groups]. Mat. Sb., 1954, vol. 34(76), no. 1, pp. 89–126. (in Russian)

8. *Denisov O. V.* Veroyatnostnye svoystva dvoichnykh otobrazheniy [Probabilistic Properties of Binary Maps]. Uchebno-metodicheskoe posobie. Moscow, 2008. (in Russian)
9. *Borovkov A. A.* Matematicheskaya statistika [Mathematical statistics]. Moscow, Nauka Publ., 1984. (in Russian)
10. *Ryabko B. Ya. and Pestunov A. I.* «Stopka knig» kak novyy statisticheskiy test dlya sluchaynykh chisel ["Book Stack" as a new statistical test for random numbers]. Probl. Peredachi Inf., 2004, vol. 40, iss. 1, pp. 73–78. (in Russian)
11. *Ryabko B. Ya., Monarev V. A., and Shokin Yu. I.* Novyy tip atak na blokovye shifry [A new type of attacks on block ciphers]. Probl. Peredachi Inf., 2005, vol. 41, iss. 4, pp. 97–107. (in Russian)
12. *Lysyak A. S., Ryabko B. Ya., and Fionov A. N.* Analiz effektivnosti gradientnoy statisticheskoy ataki na blokovye shifry RC6, MARS, CAST-128, IDEA, Blowfish v sistemakh zashchity informatsii [Efficiency analysis of gradient statistical attack on block ciphers RC6, MARS, CAST-128, IDEA, Blowfish]. Vestnik SibGUTI, 2013, no. 1, pp. 85–109. (in Russian)
13. *Pestunov A. I.* Predvaritel'naya otsenka minimal'nogo chisla raundov legkovesnykh shifrov dlya obespecheniya ikh udovletvoritel'nykh statisticheskikh svoystv [Preliminary evaluation of a minimal number of rounds in lightweight block ciphers for providing their satisfactory statistical properties]. Prikl. Diskr. Mat. Suppl., 2015, iss. 8, pp. 66–68. (in Russian)
14. *Pestunov A. I.* Teoreticheskoe issledovanie svoystv statisticheskogo testa «stopka knig» [Theoretical investigation of the Bookstack test features]. Vychislitel'nye tekhnologii, 2006, vol. 11, no. 6, pp. 96–103. (in Russian)

УДК 519.1

ОБ ОЦЕНКЕ СТОЙКОСТИ AEAD-КРИПТОСИСТЕМЫ ТИПА GCM

А. Ю. Зубов

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

Обсуждается методика доказуемой стойкости криптосистем, обеспечивающих конфиденциальность и аутентичность информации. Предлагается упрощённый вариант известной оценки доказуемой стойкости AEAD-криптосистемы GCM с вектором инициализации фиксированной длины. В тех же условиях получена оценка доказуемой стойкости модификации GCM. Приводится сравнительный анализ криптосистем.

Ключевые слова: *AEAD-криптосистема, GCM, доказуемая стойкость.*

DOI 10.17223/20710410/32/4

ON THE SECURITY OF AEAD-CRYPTOSYSTEM OF THE GCM TYPE

A. Yu. Zubov

Lomonosov Moscow State University, Moscow, Russia

E-mail: Zubovanatoly@yandex.ru

A provable security methodology for the cryptosystems ensuring information privacy and authenticity is discussed. A simplified version of the well-known estimates for the provable security of the AEAD-cryptosystem GCM with an initialization vector of fixed length is proposed. Under the same conditions an estimate for the provable security of GCM modification is obtained. A comparative analysis of the considered cryptosystems is provided.

Keywords: *AEAD-cryptosystem, GCM, provable security.*

1. Необходимые сведения

Пусть $G = \{f : S \to A\}$ — семейство функций, $|S| = k$, $|A| = n$, $n < k$, $|G| = b$. Назовём G $(b; k, n)$-*семейством функций*. Нас будет интересовать свойство функций f из G иметь коллизии, под которыми понимается совпадение $f(s_1)$ и $f(s_2)$ для различных элементов s_1, s_2 из S.

Определение 1. Пусть $A = \{0,1\}^n$; $(b; k, n)$-семейство G называется εXOR-*универсальным* (кратко — εXU-*семейством*), если для любых $s_1, s_2 \in S$ и любого $a \in A$ справедливо неравенство

$$|\{f \in G : f(s_1) \oplus f(s_2) = a\}| \leqslant \varepsilon b.$$

Система MAC типа Wegman — Carter (WC-MAC) основана на εXU-семействе функций. Известно большое число таких систем. Нам понадобятся введённые в [1] системы $WC[G]$ и $WC[G, F]$.

Определение 2. Пусть $G = \{f : S \to \{0,1\}^n\}$ — семейство функций и $P = p_1, p_2, \ldots$ — случайная равновероятная строка векторов $p_i \in \{0,1\}^n$. Пусть cnt — целочисленная переменная (счётчик). Система $WC[G]$ снабжает сообщение $s \in S$ на ключе $\{f, P\}$ меткой $\tau = (cnt, p_{cnt} \oplus f(s))$. Для проверки аутентичности сообщения (s, τ), где $\tau = (i, r)$, вычисляется сумма $p_i \oplus f(s)$, которая сравнивается с r. Совпадение — критерий аутентичности.

Определение 3. Пусть $F : \{0,1\}^k \times \{0,1\}^l \to \{0,1\}^n$ и $G = \{f : S \to \{0,1\}^n\}$ — семейства функций. Пусть cnt — целочисленная переменная (счётчик). Система $WC[G, F]$ снабжает сообщение $s \in S$ меткой $\tau(s) = (cnt, F_K(cnt) \oplus f(s))$, где cnt — текущее значение счётчика, $K \in \{0,1\}^k$, $f \in G$. Ключом служит пара (K, f). При аутентификации каждого следующего сообщения алгоритм генерации MAC увеличивает значение cnt и проверяет условие $cnt < 2^n - 1$. Если неравенство не выполняется, то изменяется значение ключа. Если выполняется, то вычисляется метка по указанной формуле. Для верификации сообщения $(s, \tau(s))$, где $\tau(s) = (i, r)$, проверяющий вычисляет сумму $F_K(i) \oplus f(s)$ и сравнивает её с r. Равенство — критерий аутентичности.

Реализуемые системами $WC[G]$ и $WC[G, F]$ семейства функций можно рассматривать как псевдослучайные семейства функций (PRF). Мерой их псевдослучайности служит свойство неотличимости случайно выбранного представителя семейства от случайной функции по входам-выходам. *Случайной* называется функция, значение которой от любого аргумента выбирается случайно и равновероятно из её области значений. Другими словами, это функция, выбранная из семейства функций $\mathrm{Rand}^{S \to A}$ всех функций $S \to A$ случайно равновероятно. В рассматриваемом случае $A = \{0,1\}^n$. Количественно указанное свойство неотличимости произвольного семейства функций $\Phi = \{f : S \to A\}$ определяется следующим образом. Проводится атака различения, в которой принимают участие *различитель* и *оракул*. Различитель $\mathbf{A}$ (вероятностный алгоритм) может обращаться к оракулу O^{φ} функции $\varphi : S \to A$ с запросами относительно выбранных аргументов x из S, получая от оракула значения $\varphi(x)$. На основании полученных ответов $\mathbf{A}$ решает, какая гипотеза имеет место:

— W_0 : функция φ выбрана случайно равновероятно из $\mathrm{Rand}^{S \to A}$;
— W_1 : функция φ выбрана случайно равновероятно из Φ.

Выбор W_0 или W_1 произведён заранее, до начала обмена вопросами и ответами.

Результатом работы алгоритма $\mathbf{A}$ является различающий бит d, равный 0, если $\mathbf{A}$ делает выбор в пользу W_0, и 1 в противном случае. Эффективность работы различителя $\mathbf{A}$ определяется величиной $\mathrm{Adv}^{\mathrm{prf}}_{\Phi,\mathbf{A}}$, называемой *prf-преимуществом* семейства Φ для различителя $\mathbf{A}$. Она определяется формулой

$$\mathrm{Adv}^{\mathrm{prf}}_{\Phi,\mathbf{A}} = \left|\mathsf{P}\left[d = 1 | W_0\right] - \mathsf{P}\left[d = 1 | W_1\right]\right|, \tag{1}$$

представляющей собой модуль разности условных вероятностей того, что различающий бит равен 1, при условии, что имеет место гипотеза W_0 или W_1.

При теоретико-информационном подходе затраты различителя не учитываются.

Определение 4. Семейство функций Φ называется *ε-псевдослучайным*, если величина

$$\mathrm{Adv}^{\mathrm{prf}}_{\Phi} = \max_{\mathbf{A}} \left\{\mathrm{Adv}^{\mathrm{prf}}_{\Phi,\mathbf{A}}\right\},$$

в которой максимум берётся по всем различителям $\mathbf{A}$, удовлетворяет неравенству $\mathrm{Adv}^{\mathrm{prf}}_{\Phi} \leqslant \varepsilon$. Величина $\mathrm{Adv}^{\mathrm{prf}}_{\Phi}$ называется *prf-преимуществом семейства* Φ.

При теоретико-сложностном подходе величина prf-преимущества зависит от объёма затрат. К ним относятся число запросов к оракулу, их общий объём и время работы, под которым понимается действительное время выполнения алгоритма.

Определение 5. Для любых $t, q, \mu \in \mathbb{N}$ prf-преимущество семейства Φ — это максимум

$$\mathrm{Adv}_{\Phi}^{\mathrm{prf}}(t, q, \mu) = \max_{\mathbf{A}} \left\{ \mathrm{Adv}_{\Phi, \mathbf{A}}^{\mathrm{prf}} \right\}$$

по всем различителям $\mathbf{A}$, имеющим временную сложность t и использующим не более q запросов к оракулу, суммарная длина которых не превосходит μ битов. Φ называется $\varepsilon(t, q, \mu)$-*псевдослучайным*, если выполняется неравенство $\mathrm{Adv}_{\Phi}^{\mathrm{prf}}(t, q, \mu) \leqslant \varepsilon$. Чем меньше ε (при допустимом объёме затрат), тем «более псевдослучайным» является семейство Φ (и тем более стойким к атаке различения).

Аналогично определяется *prp-преимущество* $\mathrm{Adv}_{\Phi}^{\mathrm{prp}}$ семейства подстановок Φ.

Замечание 1. Prf-преимущество семейства функций определяется как prf-преимущество «лучшего» различителя. Если для некоторого различителя выражение под модулем в (1) отрицательно, то для различителя, который принимает противоположный различающий бит, это выражение будет положительным. Второй различитель «лучше» первого. Поэтому, оценивая преимущество семейства функций, можно опускать модуль в определении (1), что обычно и делается.

Замечание 2. Различитель характеризуется набором вероятностей принять различающий бит 1 для каждой последовательности запросов $\overline{x}$ и последовательности $\overline{y}$ ответов на них. Для семейства Rand всех функций для каждого $\overline{x}$ допустим какой угодно $\overline{y}$. «Хороший» различитель семейства Φ ограничивает для каждого $\overline{x}$ число вариантов $\overline{y}$, для которых указанная вероятность отлична от нуля. Поэтому гипотезам W_1 и W_0 отвечают разные распределения вероятностей, P_U и P_V, принять различающий бит 1 на множестве M возможных пар $(\overline{x}, \overline{y})$. Пусть для «лучшего» различителя такими распределениями $\mathsf{P}_U, \mathsf{P}_V$ на (конечном) множестве M определены случайные переменные U и V. Тогда, как показано, например, в [2], преимущество «лучшего» различителя можно записать в виде

$$\max_{f: M \to \{0,1\}} \left| \mathsf{P}\left[f(U) = 1\right] - \mathsf{P}\left[f(V) = 1\right] \right| = 0{,}5 \sum_{m \in M} \left| \mathsf{P}_U(m) - \mathsf{P}_V(m) \right|,$$

где максимум берётся по всем отображениям из M в $\{0, 1\}$. Правая часть этого равенства называется *расстоянием по вариации* (или *статистическим расстоянием*) между распределениями $\mathsf{P}_U, \mathsf{P}_V$. Это равенство даёт другое представление о понятии преимущества.

Определение 6. Система MAC называется ε-*стойкой*, $0 < \varepsilon < 1$, если противник, наблюдая ряд аутентифицированных (на одном ключе, выбранном случайно равновероятно) сообщений, может составить новое аутентифицированное (на том же ключе) сообщение с вероятностью, не превосходящей ε.

Утверждение 1. Пусть G — это εXU-семейство функций. Тогда система MAC $WC[G]$ является ε-стойкой.

Доказательство. Пусть противник имеет сообщения $(s_1, \tau_1), \ldots, (s_q, \tau_q)$, где τ_i — метки, полученные системой $WC[G]$ при использовании ключа (f, P). Оценим вероятность того, что противник сможет составить новое сообщение s^* и снабдить его меткой τ^*, полученной на том же ключе для подходящего значения счётчика — i^*.

Противник обязан произвести значение счётчика — i^*. При этом имеется две возможности: $i^* \leqslant q$ и $i^* > q$. Если $i^* > q$, то p_{i^*} — случайный вектор, не коррелирующий с наблюдаемыми значениями. Поэтому сумма $p_{i^*} \oplus f(s^*) = \tau^*$ также представляет собой вектор, выбранный случайно. Значение τ^* противник может лишь угадать с вероятностью 2^{-n}. Пусть $i^* \leqslant q$. Это означает, что противник выбирает значение счётчика, использованное ранее при вычислении метки τ_{i^*} сообщения s_{i^*}. Теперь ему нужно произвести $s^* \neq s_{i^*}$ и τ^*, такие, что

$$\tau^* = f\left(s^*\right) \oplus p_{i^*} = f\left(s_{i^*}\right) \oplus p_{i^*} = \tau_{i^*}.$$

Из этих соотношений получаем равенство $f(s^*) \oplus f(s_{i^*}) = \tau^* \oplus \tau_{i^*}$.

Пусть $c = \tau^* \oplus \tau_{i^*}$. Поскольку G образует εXU-семейство, имеем неравенство

$$\mathsf{P}\left[f\left(s^*\right) \oplus f\left(s_{i^*}\right) = c\right] \leqslant \varepsilon.$$

Таким образом, и искомая вероятность не превосходит ε. ■

Усилим утверждение 1, предоставив противнику возможность аутентифицировать адаптивно подобранные сообщения и проверить корректность их меток.

Утверждение 2. Пусть G — это εXU-семейство функций. Тогда при использовании не более чем q запросов к оракулу проверки метки и к оракулу генерации метки система $WC[G]$ является εq-стойкой.

Доказательство. В серии подделок (s_i, τ_i), $i \geqslant 1$, успех в атаке определяется путём запросов к оракулу проверки метки. Перед каждым таким запросом допускаются запросы к оракулу генерации метки относительно подобранных сообщений.

Пусть U_i — событие, означающее неуспех в подделке при i-м запросе к оракулу проверки метки. Если имеет место событие $U_1 \cap \ldots \cap U_i$, то, согласно утверждению 1, вероятность успеха следующей подделки ограничена ε. Поэтому имеет место неравенство

$$\mathsf{P}\left[U_{i+1} \middle| \left(U_1 \cap \ldots \cap U_i\right)\right] \geqslant 1 - \varepsilon.$$

Поскольку

$$\mathsf{P}\left[U_1 \cap \ldots \cap U_l\right] = \mathsf{P}\left[U_l \middle| \left(U_1 \cap \ldots \cap U_{l-1}\right)\right] \cdot \mathsf{P}\left[U_{l-1} \middle| \left(U_1 \cap \ldots \cap U_{l-2}\right)\right] \cdot \ldots \cdot \mathsf{P}\left[U_1\right]$$

и каждый сомножитель в правой части не меньше $1 - \varepsilon$, получаем неравенство

$$\mathsf{P}\left[U_1 \cap \ldots \cap U_q\right] \geqslant (1 - \varepsilon)^q.$$

Таким образом, вероятность успеха атаки не превосходит величины

$$1 - (1 - \varepsilon)^q \leqslant \varepsilon q.$$

Утверждение доказано. ■

2. AEAD-криптосистемы, GCM

В последние годы уделяется большое внимание построению и исследованию новых режимов блочного шифрования, обеспечивающих конфиденциальность и аутентификацию данных [3–7]. В связи с этим возникли новые термины, соответствующие названиям новых режимов шифрования: *Authenticated Encryption* (кратко — AE) и *Authenticated Encryption with Associated Data* (AEAD). В AEAD-системах (в отличие от AE-систем) используются так называемые *ассоциированные данные* (*associated*

data). Это — дополнительные сведения, которые при передаче должны быть аутентифицированы, но не зашифрованы. Пример — заголовок сетевого пакета, по которому можно определить характер информации, содержащейся в зашифрованном пакете.

AEAD-криптосистема GCM основана на блочном шифровании в режиме счётчика (CTR) и системе аутентификации типа $WC[G, F]$, разработана в 2004 г. [8]. Система GCM стандартизована NIST в документе SP800-38D, имеет доказуемую стойкость, эффективную программную и аппаратную реализацию, широко используется в популярных криптографических протоколах IPSec, MACSec, P1619.1, TLS.

Алгоритм GCM имеет четыре входа:

— ключ K;
— вектор инициализации iv длины до 2^{64} битов (рекомендуется iv длины 96 битов);
— открытый текст P длины до $(2^{32} - 2)\, n$ битов (n — длина блока шифрования);
— ассоциированные данные A длины до $2^{n/2}$ битов.

Алгоритм GCM имеет два выхода:

— шифртекст C, длина которого совпадает с длиной открытого текста;
— метка аутентификации T длины до n битов; длина метки обозначается τ.

Вектор iv должен меняться с каждым открытым текстом. Не требуется, чтобы iv был случайным или непредсказуемым. Он передаётся вместе с шифртекстом. Открытый текст и ассоциированные данные разбиваются на n-битовые блоки:

$$P = P_1||\ldots||P_{m-1}||P_m^*, \quad A = A_1||\ldots||A_{r-1}||A_r^*.$$

Блоки P_m^* и A_r^* могут быть неполными. Шифртекст имеет вид $C = C_1||\ldots||C_{m-1}||C_m^*$, где длина блока $l\,(P_m^*) = l\,(C_m^*)$. Пусть

$$l(P) = (m-1)n + u,\ 1 \leqslant u \leqslant n; \quad l(A) = (r-1)n + v,\ 1 \leqslant v \leqslant n.$$

Схема алгоритма GCM изображена на рис. 1.

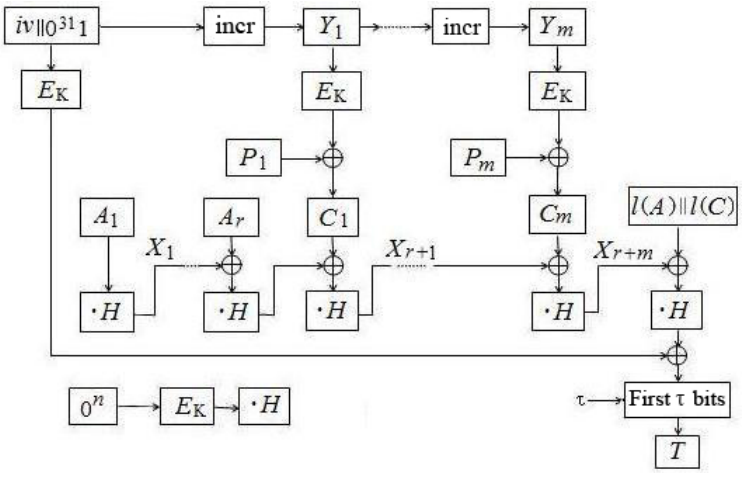

Рис. 1. Схема алгоритма GCM

Алгоритм использует счётчиковую последовательность $Y = Y_0||Y_1||\ldots$, в которой $Y_i = \mathrm{incr}\,(Y_{i-1})$. Функция $\mathrm{incr}(Y)$ изменяет младшие 32 бита Y, добавляя единицу по модулю 2^{32} :

$$\mathrm{incr}\,(L||R) = L||\,(R+1) \bmod 2^{32}.$$

Алгоритм GCM состоит в выполнении следующих операций:

— $H = E_K(0^n)$;

— $Y_0 = \begin{cases} iv||0^{31}1, & \text{если } l(iv) = n - 32, \\ GHASH(H, \{\}, iv) & \text{в противном случае;} \end{cases}$

— $Y_i = \operatorname{incr}(Y_{i-1}),\ i = 1, \ldots, m$;

— $C_i = P_i \oplus E_K(Y_i),\ i = 1, \ldots m - 1$;

— $C_m^* = P_m^* \oplus MSB_u(E_K(Y_m))$;

— $T = MSB_\tau(GHASH(H, A, C) \oplus E_K(Y_0))$.

В описании алгоритма $MSB_t(Z)$ обозначает старшие t битов вектора Z, $\{\}$ — пустую строку, $E_K(Z)$ — результат зашифрования Z на ключе K. Зашифрованное сообщение передаётся в виде вектора (iv, A, C, T). Функция $GHASH$ определяется формулой

$$GHASH(H, A, C) = X_{m+r+1},$$

в которой $X_i,\ i = 0, 1, \ldots, m + r + 1$, задаются рекуррентным соотношением

$$X_i = \begin{cases} 0, & \text{если } i = 0, \\ (X_{i-1} \oplus A_i) \cdot H, & \text{если } i = 1, \ldots, r - 1, \\ (X_{r-1} \oplus (A_r^*||0^{n-v})) \cdot H, & \text{если } i = r, \\ (X_{i-1} \oplus C_{i-r}) \cdot H, & \text{если } i = r + 1, \ldots, m + r - 1, \\ (X_{m+r-1} \oplus (C_m^*||0^{n-u})) \cdot H, & \text{если } i = m + r, \\ (X_{m+r} \oplus (l(A) \oplus l(C))) \cdot H, & \text{если } i = m + r + 1, \end{cases} \tag{2}$$

где операции сложения и умножения выполняются в поле $\mathrm{GF}(2^{128})$, порождённом над полем $\mathrm{GF}(2)$ неприводимым многочленом $x^{128} + x^7 + x^2 + x + 1$.

3. Стойкость GCM

Оценка стойкости GCM, как и других AE-систем, проводится с позиции *доказуемой стойкости* (*provable security*). Известна иерархия понятий доказуемой стойкости, отражающих как статистический (или теоретико-информационный), так и теоретико-сложностной подходы [3, 6, 9, 10]. Второй подход предпочтительнее, поскольку позволяет регулировать выбор параметров криптосистем, обеспечивающих требуемый уровень стойкости. В связи с такой возможностью используется термин *практически-ориентированная доказуемая стойкость* (*practice-oriented provable security*) [9].

В 2004 г. было показано [10], что для AE-систем оценка стойкости типа (IND-CPA)+(AUTH) является адекватной мерой стойкости в смысле различных понятий доказуемой стойкости. Понятие (IND-CPA)+(AUTH)-стойкости означает следующее.

Стойкость шифрования системы оценивается стойкостью к атаке различения на основе подобранного открытого текста. Стойкость аутентификации системы оценивается стойкостью к активной атаке на основе подобранного открытого текста, в которой атакующий имеет доступ к оракулу генерации метки и оракулу проверки метки. Запросы к оракулам можно как угодно чередовать по усмотрению атакующего. После серии запросов и получения ответов на них атакующий вырабатывает «подделку» — пару текст — метка, в которой текст отличен от всех текстов, используемых в запросах. Атака достигает успеха, если метка признаётся корректной. Мерой стойкости служат значения функции преимущества, введённой в определениях 4 и 5 и прокомментированной в замечаниях 1 и 2. В [8] для оценки стойкости GCM использован именно такой

подход. Приведём доказательство оценки стойкости GCM в случае, когда $l(iv) = n - t$, где 2^t — максимальное число сообщений, которые можно зашифровать на одном ключе (для GCM $n = 128$, $l(iv) = 128 - 32$). В этом случае фактор наличия коллизий функций семейства $GHASH$ можно не учитывать, что упрощает получение оценки. Этот упрощённый вариант доказательства мы приводим с целью иллюстрации методики.

Теорема 1. Пусть $\mathcal{C}$ — противник, имеющий преимущество $\mathcal{C}_{\mathrm{GCM}}$ в атаке различения семейства функций, реализуемого GCM, или в активной атаке против GCM, при числе запросов к оракулам, не превосходящем q. Пусть для каждого запроса (iv, A, P) выполняются условия $l(C) + l(A) < l$ и $l(iv) = n - t$. Тогда существует различитель $\mathcal{B}$ базового шифра E, имеющий преимущество $\mathcal{B}_E$, где

$$\mathcal{C}_{\mathrm{GCM}} \leqslant \mathcal{B}_E + q \left\lceil \frac{l}{n} + 1 \right\rceil 2^{-\tau} + \frac{q(q-1)}{2^{n+1}}.$$

Доказательство. Оценим сначала стойкость аутентификации.

Рассмотрим базовый шифр E как псевдослучайное семейство функций (PRF). Пусть $\mathcal{C}$ — противник, имеющий преимущество $\mathcal{C}_{\mathrm{GCM}}$ в активной атаке против GCM. Используем $\mathcal{C}$ для построения различителя $\mathcal{B}$ семейства E. В получении ответов на свои запросы $\mathcal{C}$ будет использовать $\mathcal{B}$, направляя ему значения счётчиковой последовательности и получая от него необходимые результаты зашифрования. В свою очередь, $\mathcal{B}$ будет получать эти ответы от оракула зашифрования.

Пусть W_1 и W_0 — события, соответствующие использованию в атаке различения для $\mathcal{B}$ семейства функций E или случайной функции. После серии запросов $\mathcal{C}$ строит подделку. Если $\mathcal{C}$ достигает успеха в подделке, то $\mathcal{B}$ делает выбор в пользу W_1, иначе — в пользу W_0.

Выразим преимущество $\mathcal{B}_{\mathrm{PRF}} = \mathrm{Adv}^{\mathrm{prf}}_{E,\mathcal{B}}$ различителя $\mathcal{B}$ через преимущество $\mathcal{C}_{\mathrm{GCM}} = \mathrm{Adv}^{\mathrm{prf}}_{\mathrm{GCM},\mathcal{C}}$ противника $\mathcal{C}$. Согласно (1) и замечанию 1,

$$\mathcal{B}_{\mathrm{PRF}} = \mathsf{P}\left[d^{\mathcal{B}} = 1 | W_1\right] - \mathsf{P}\left[d^{\mathcal{B}} = 1 | W_0\right], \tag{3}$$

где $d^{\mathcal{B}}$ — различающий бит, вырабатываемый $\mathcal{B}$. Из описания действий $\mathcal{B}$ и определения преимущества $\mathcal{C}_{\mathrm{GCM}}$ следует, что

$$\mathcal{C}_{\mathrm{GCM}} = \mathsf{P}\left[d^{\mathcal{B}} = 1 | W_1\right]. \tag{4}$$

Оценим вероятность $\mathsf{P}\left[d^{\mathcal{B}} = 1 | W_0\right]$.

Лемма 1. $G = GHASH$ образует $\lceil l/n + 1 \rceil 2^{-\tau} XU$-семейство функций, где τ — длина метки и l определяется неравенством $l(A) + l(C) \leqslant l$.

Доказательство. Напомним, что $GHASH\,(H, A, C) = X_{m+r+1}$, где X_i определяется формулой (2). Рассмотрим два различных входа (A, C), (A', C') и оценим вероятность события

$$G(H, A, C) \oplus G(H, A', C') = a||z, \tag{5}$$

где a — фиксированная τ-битовая строка; z — $(n - \tau)$-битовая переменная (отсекаемая часть выхода).

Согласно описанию алгоритма GCM, строки A, C, A', C' разбиваются соответственно на r, m, r', m' n-битовых блоков. Последние блоки имеют длины v, u, v', u' соответственно. Пусть $h = \max\{m + r, m' + r'\}$ — число блоков в более длинном входе.

Определим блоки:

$$D_i = \begin{cases} A_i, & \text{если } i = 1, \ldots, r-1, \\ A_r^* || 0^{n-v}, & \text{если } i = r, \\ C_{i-r}, & \text{если } i = r+1, \ldots, r+m-1, \\ C_m^* || 0^{n-u}, & \text{если } i = r+m, \\ l(A) || l(C), & \text{если } i = r+m+1, \\ 0^n, & \text{если } i = r+m+2, \ldots, h+1. \end{cases}$$

Аналогично определим блоки D'_i для пары (A', C').

Равенство (5) можно записать в виде $R(H) = 0$, где R — многочлен степени, не превосходящей $h+1$, над полем $\mathrm{GF}(2^n)$:

$$R(H) = (a||z) \oplus \sum_{i=1}^{h} (D_i \oplus D'_i) \cdot H^i.$$

Поскольку $A||C \neq A'||C'$, многочлен R — ненулевой и, следовательно, имеет не более $h+1$ корней. Если H выбирать случайно и равновероятно из $\mathrm{GF}(2^n)$, то вероятность того, что $R(H) = 0$, не превосходит $(h+1)2^{-n} \leqslant \lceil l/n + 1 \rceil 2^{-n}$ при условии, что суммарная длина входов ограничена l битами.

Нетрудно видеть, что имеется взаимно-однозначное соответствие между блоками D и парами (A, C). Поэтому вероятность того, что $R(H) = 0$, для любых двух данных пар (A, C), (A', C') и данного вектора $a||z$ равна вероятности события (5). А так как имеется $2^{n-\tau}$ различных значений z, событие (5) выполняется с вероятностью, не превосходящей

$$\lceil l/n + 1 \rceil 2^{-n} 2^{n-\tau} = \lceil l/n + 1 \rceil 2^{-\tau}$$

для любых пар $(A, C), (A', C')$ и $a \in \{0,1\}^\tau$, что и требуется. ■

Вернёмся к доказательству теоремы. Заметим, что при условии W_0 система MAC типа $WC[G, E]$ модифицируется в систему MAC типа $WC[G]$, поскольку E заменяется случайной функцией. Такая система MAC, согласно утверждению 1 и лемме 1, является $\lceil l/n + 1 \rceil 2^{-\tau}$-стойкой.

Из утверждения 2 получаем неравенство

$$\mathsf{P}\left[d^{\mathcal{B}} = 1 | W_0\right] \leqslant q \lceil l/n + 1 \rceil 2^{-\tau}. \tag{6}$$

Следующее утверждение, известное под названием PRF-PRP-switching lemma [11], связывает $\mathcal{B}_{\mathrm{PRF}}$ и $\mathcal{B}_{\mathrm{PRP}} = \mathcal{B}_E$.

Лемма 2. Пусть $\mathcal{A}_{\mathrm{PRF}}(\mathcal{A}_{\mathrm{PRP}})$ — преимущество различителя блочного шифра с n-битовым блоком и случайной функции (случайной подстановки). Пусть различитель использует не более чем q запросов к оракулу. Тогда

$$\mathcal{A}_{\mathrm{PRF}} \leqslant \mathcal{A}_{\mathrm{PRP}} + \frac{q(q-1)}{2^{n+1}}. \tag{7}$$

Утверждение теоремы о стойкости аутентификации следует из (2), (4), (6) и (7):

$$\mathcal{C}_{\mathrm{GCM}} = \mathcal{B}_{\mathrm{PRP}} + \mathsf{P}\left[d^{\mathcal{B}} = 1 | W_0\right] \leqslant \mathcal{B}_{\mathrm{PRP}} + q \lceil l/n + 1 \rceil 2^{-\tau} \leqslant$$
$$\leqslant \mathcal{B}_E + q \lceil l/n + 1 \rceil 2^{-\tau} + \frac{q(q-1)}{2^{n+1}}.$$

Получим теперь оценку стойкости шифрования схемы. Пусть $\mathcal{C}$ — различитель семейства функций, реализуемых GCM, имеющий преимущество $\mathcal{C}_{\mathrm{GCM}}$. Построим различитель $\mathcal{B}$ семейства функций E. Для получения ответов на свои запросы $\mathcal{C}$ будет использовать $\mathcal{B}$, направляя ему соответствующие блоки счётчика и получая от него необходимые результаты зашифрования. В свою очередь, $\mathcal{B}$ будет получать эти результаты от оракула зашифрования. Если после серии запросов $\mathcal{C}$ принимает бит 1, то и $\mathcal{B}$ принимает бит 1.

Пусть W_1 и W_0 — события, соответствующие использованию в атаке различения для $\mathcal{B}$ семейства функций E и случайной функции, а U_1 и U_0 — события, соответствующие использованию в атаке различения для $\mathcal{C}$ семейства функций GCM и случайной функции. Согласно (1) и замечанию 1, преимущество $\mathcal{C}$ выражается формулой

$$\mathcal{C}_{\mathrm{GCM}} = \mathsf{P}\left[d^{\mathcal{C}} = 1|U_1\right] - \mathsf{P}\left[d^{\mathcal{C}} = 1|U_0\right]. \tag{8}$$

В свою очередь, преимущество $\mathcal{B}$ выражается формулой (3).

Из описания действий различителя $\mathcal{C}$ следует, что

$$\mathsf{P}\left[d^{\mathcal{C}} = 1|U_1\right] = \mathsf{P}\left[d^{\mathcal{B}} = 1|W_1\right]. \tag{9}$$

Оценим вероятность $\mathsf{P}\left[d^{\mathcal{C}} = 1|U_0\right]$.

Каждый запрос различителя использует счётчиковую последовательность $Y_0, Y_1, \ldots$ Обозначим через Q событие, означающее, что значения счётчика во всех запросах различны, а через R — событие, означающее, что $\mathcal{C}$ не достигает успеха.

Лемма 3. Справедливо неравенство

$$\mathsf{P}\left[d^{\mathcal{B}} = 1|\,(W_0 \cap Q)\right] \cdot \mathsf{P}\left[Q|W_0\right] \leqslant \mathsf{P}\left[d^{\mathcal{C}} = 1|U_0\right]. \tag{10}$$

Доказательство. Событие $W_0 \cap Q$ означает, что в атаке реализуется система S, которая получается из GCM заменой функции шифрования случайной функцией, и счётчиковые последовательности в запросах $\mathcal{C}$ не имеют коллизий. Поскольку выход S равносилен выходу случайной функции, событие $\left(d^{\mathcal{C}} = 1|\,(U_0 \cap Q)\right)$ означает, что $\mathcal{C}$ принимает случайную функцию за GCM. Отсюда и из того, что $\mathsf{P}\left[U_0\right] = \mathsf{P}\left[W_0\right] = 0{,}5$, получаем следующие соотношения:

$$\mathsf{P}\left[d^{\mathcal{B}} = 1|\,(W_0 \cap Q)\right] \cdot \mathsf{P}\left[Q|W_0\right] = \frac{\mathsf{P}\left[\left(d^{\mathcal{B}} = 1\right) \cap W_0 \cap Q\right] \cdot \mathsf{P}\left[W_0 \cap Q\right]}{\mathsf{P}\left[W_0 \cap Q\right] \cdot \mathsf{P}\left[W_0\right]} =$$

$$= \frac{\mathsf{P}\left[\left(d^{\mathcal{B}} = 1\right) \cap W_0 \cap Q\right]}{\mathsf{P}\left[W_0\right]} = \frac{\mathsf{P}\left[\left(d^{\mathcal{C}} = 1\right) \cap U_0 \cap Q\right]}{\mathsf{P}\left[U_0\right]} = \mathsf{P}\left[\left(d^{\mathcal{C}} = 1\right) \cap Q|U_0\right] \leqslant \mathsf{P}\left[d^{\mathcal{C}} = 1|U_0\right].$$

Отсюда следует неравенство (10). ■

Выразим $\mathcal{B}_{\mathrm{PRP}}$ через $\mathcal{C}_{\mathrm{GCM}}$. С учётом (8)–(10) имеем

$$\mathcal{B}_{\mathrm{PRP}} = \mathsf{P}\left[d^{\mathcal{C}} = 1|U_1\right] - \mathsf{P}\left[d^{\mathcal{B}} = 1|W_0\right] =$$

$$= \mathsf{P}\left[d^{\mathcal{C}} = 1|U_1\right] - \mathsf{P}\left[\left(d^{\mathcal{B}} = 1\right) |\,(W_0 \cap Q)\right] \cdot \mathsf{P}\left[Q|W_0\right] - \mathsf{P}\left[\left(d^{\mathcal{B}} = 1\right) |\,\left(W_0 \cap \overline{Q}\right)\right] \cdot \mathsf{P}\left[\overline{Q}|W_0\right] \geqslant$$

$$\geqslant \mathsf{P}\left[d^{\mathcal{C}} = 1|U_1\right] - \mathsf{P}\left[d^{\mathcal{C}} = 1|U_0\right] - \mathsf{P}\left[\left(d^{\mathcal{B}} = 1\right) |\,\left(W_0 \cap \overline{Q}\right)\right] \cdot \mathsf{P}\left[\overline{Q}|W_0\right] =$$

$$= \mathcal{C}_{\mathrm{GCM}} - \mathsf{P}\left[\left(d^{\mathcal{B}} = 1\right) |\,\left(W_0 \cap \overline{Q}\right)\right] \cdot \mathsf{P}\left[\overline{Q}|W_0\right] \geqslant \mathcal{C}_{\mathrm{GCM}} - \mathsf{P}\left[\overline{Q}|W_0\right] \geqslant$$

$$\geqslant \mathcal{C}_{\mathrm{GCM}} - \mathsf{P}\left[\overline{Q}|W_0\right] - \mathsf{P}\left[\overline{R}|W_0\right].$$

Поскольку условия исключают возможность коллизий в счётчиковых последовательностях, событие $\overline{Q}$ невозможно, и из последних соотношений получаем неравенство

$$\mathcal{B}_{\mathrm{PRP}} \geqslant \mathcal{C}_{\mathrm{GCM}} - \mathsf{P}\left[\overline{R}|W_0\right]. \tag{11}$$

Заметим, что в условиях W_0, $l(iv) = n - t$ работает система S, для которой можно использовать утверждение 2, согласно которому вероятность $\mathsf{P}\left[\overline{R}|W_0\right]$ не превосходит εq, а ε оценивается в лемме 1. С учётом этого, из (11) и леммы 2 получаем нужную оценку: $\mathcal{C}_{\mathrm{GCM}} \leqslant \mathcal{B}_{\mathrm{PRP}} + q\left\lceil l/n+1\right\rceil 2^{-\tau} \leqslant \mathcal{B}_E + q\left\lceil l/n+1\right\rceil 2^{-\tau} + \dfrac{q(q-1)}{2^{n+1}}$.

Теорема доказана. ■

Приведём числовой пример, иллюстрирующий полученную оценку. Базовым алгоритмом шифрования для GCM является AES. При этом $l(iv) = 96$, $\tau = 96$, $n = 128$. Пусть размер сообщения не превышает 1500 байтов (т. е. $l \leqslant 12000$). Тогда из теоремы 1 следует, что если AES неотличим от случайной подстановки при защите не более чем 2^{32} сообщений на одном ключе (т. е. при этом $\mathcal{B}_E$ практически равно 0), то преимущество атакующего AES-GCM не превосходит $5{,}17 \cdot 10^{-18}$.

4. Оценка стойкости модификации GCM

Рассмотрим модификацию GCM, в которой система MAC типа $WC[G, F]$ заменяется другой системой WC-MAC. В таких системах вместо εXU-семейства G может использоваться семейство функций с более слабым требованием к свойству иметь коллизии.

Определение 7. $(b; k, n)$-семейство функций G называется *ε-универсальным* (εU-семейством), если для любых s_1, s_2 из S, $s_1 \neq s_2$, и положительного числа ε справедливо неравенство $|\{f \in G : f(s_1) = f(s_1)\}| \leqslant \varepsilon b$.

Определение 8. Пусть $G = \left\{f : S \to \{0,1\}^l\right\}$ и $R = \mathrm{Rand}^{l \to n}$. Ключом системы MAC $FH[G]$ служит пара (f, ρ), $f \in G$, $\rho \in R$, а меткой сообщения $s \in S$ — значение $\rho(f(s))$.

Определение 9. Пусть $G = \left\{f : S \to \{0,1\}^l\right\}$ и $F : \{0,1\}^r \times \{0,1\}^l \to \{0,1\}^n$ — семейства функций. Ключ системы $FH[G, F]$ — пара (f, F_K), $f \in G$, $K \in \{0,1\}^r$. Меткой сообщения $s \in S$ служит $F_K(f(s))$.

Утверждение 3. Пусть G — это εU-семейство функций и q — максимальное число сообщений, которые можно аутентифицировать на одном ключе. Тогда система MAC $FH[G]$ является $\varepsilon q(q+1)/2$-стойкой.

Доказательство. Пусть противник наблюдает сообщения $(s_1, \tau_1), \ldots, (s_q, \tau_q)$, где τ_i — метки, полученные с помощью $FH[G]$ при использовании ключа (f, ρ). Оценим вероятность того, что противник сможет снабдить новое сообщение s^* корректной меткой τ^*.

Пусть $\Phi = \{f(s_1), \ldots, f(s_q)\}$, $T = \{\tau_1, \ldots, \tau_q\}$. Если $f(s^*) \notin \Phi$, то $\rho(f(s^*))$ с равной вероятностью может принимать любое из 2^n значений (так как ρ — случайная функция). Поэтому вероятность того, что для некоторого τ^* выполняется равенство $\tau^* = \rho(f(s^*))$, равна 2^{-n}. Таким образом, вероятность $p_{\mathrm{усп}}$ успеха атаки равна

$$p_{\mathrm{усп}} = 2^{-n}. \tag{12}$$

Рассмотрим случай, когда $f(s^*) \in \Phi$. В этом случае τ^* должно принадлежать T. Тогда в рассматриваемых условиях

$$p_{\text{усп}} = \mathsf{P}\Big[\bigcup_{(i,j)} (\tau^* = \tau_j | f(s^*) = f(s_i))\Big] = \sum_{i \leqslant j} \mathsf{P}\left[\tau^* = \tau_j | f(s^*) = f(s_i)\right].$$

В свою очередь,

$$\mathsf{P}[\tau^* = \tau_j | f(s^*) = f(s_i)] = \mathsf{P}[\tau^* = \tau_j | f(s^*) = f(s_i), f(s_i) = f(s_j)] \cdot \mathsf{P}[f(s^*) = f(s_i) = f(s_j)] + \\ + \mathsf{P}\left[\tau^* = \tau_j | f(s^*) = f(s_i), f(s_i) \neq f(s_j)\right] \cdot \mathsf{P}\left[f(s^*) = f(s_i) \neq f(s_j)\right].$$

Поскольку

$$\mathsf{P}\left[\tau^* = \tau_j | f(s^*) = f(s_i), f(s_i) = f(s_j)\right] = 1,$$
$$\mathsf{P}\left[\tau^* = \tau_j | f(s^*) = f(s_i), f(s_i) \neq f(s_j)\right] = 0,$$

получаем формулу

$$p_{\text{усп}} = \sum_{i \leqslant j} \mathsf{P}\left[f(s^*) = f(s_i) = f(s_j)\right]. \tag{13}$$

Пусть D_i — событие, означающее первое появление коллизии функции f в последовательности аргументов $s_1, \ldots, s_{i-1}, s^*$ на i-м месте. Поскольку f выбирается случайно из εU-семейства функций, из (13) получаем соотношения

$$p_{\text{усп}} \leqslant \sum_{i=1}^{q+1} \mathsf{P}\left[D_i\right] \leqslant \sum_{i=1}^{q+1} (i-1)\,\varepsilon = \varepsilon \sum_{i=0}^{q} i = \varepsilon q(q+1)/2.$$

Поскольку $2^{-n} \leqslant \varepsilon \leqslant q(q+1)/2$, в любом случае из (12) и (13) получаем оценку $p_{\text{усп}} \leqslant \varepsilon q(q+1)/2$. ■

Усилим утверждение 3, предоставив противнику возможность аутентифицировать адаптивно подобранные сообщения и проверять корректность их меток. Эти условия могут повысить шансы противника на успех. Покажем, что на самом деле указанные преимущества не позволяют улучшить оценку стойкости, полученную в утверждении 3.

Утверждение 4. Пусть G — это εU-семейство функций. Тогда при использовании не более чем q запросов к оракулу проверки метки и к оракулу генерации метки система $FH[G]$ является $\varepsilon q(q-1)/2$-стойкой.

Доказательство. В серии подделок (s_i, τ_i), $i = 1, 2, \ldots, q$, успех в атаке определяется путём запросов к оракулу проверки метки. Пусть $q = q_1 + q_2$, где q_1 — число запросов к оракулу генерации метки; q_2 — число запросов к оракулу проверки метки.

Пусть U_i — событие, означающее неуспех при i-м запросе к оракулу проверки метки, и p_i — число запросов к оракулу генерации метки, произведённое перед i-м запросом к оракулу проверки метки. Если имеет место событие $U_1 \cap \ldots \cap U_i$, то, согласно утверждению 3, вероятность успеха в следующей подделке ограничена числом $\varepsilon p_{i+1}(p_{i+1}+1)/2$. Это означает, что имеет место неравенство

$$\mathsf{P}\left[U_{i+1} | (U_1 \cap \ldots \cap U_i)\right] \geqslant 1 - \frac{\varepsilon p_{i+1}(p_{i+1}+1)}{2}.$$

Поскольку

$$\mathsf{P}\left[U_1 \cap \ldots \cap U_i\right] = \mathsf{P}\left[U_i | U_1 \cap \ldots \cap U_{i-1}\right] \cdot \mathsf{P}\left[U_{i-1} | U_1 \cap \ldots \cap U_{i-2}\right] \cdot \ldots \cdot \mathsf{P}\left[U_1\right],$$

получаем неравенство

$$\mathsf{P}\left[U_1 \cap \ldots \cap U_{q_2}\right] \geqslant \left(1 - \frac{\varepsilon p_1 (p_1+1)}{2}\right)\left(1 - \frac{\varepsilon p_2 (p_2+1)}{2}\right)\ldots\left(1 - \frac{\varepsilon p_{q_2} (p_{q_2}+1)}{2}\right).$$

Поскольку $p_{i+1} \geqslant p_i$ и $p_{q_2} \leqslant q_1$, вероятность $1 - \mathsf{P}\left[U_1 \cap \ldots \cap U_{q_2}\right]$ успеха атаки не превосходит величины

$$1 - \left(1 - \frac{\varepsilon q_1 (q_1+1)}{2}\right)^{q_2},$$

которая, в свою очередь, меньше $\varepsilon q_1(q_1+1)q_2/2$.

Так как $q_1 + q_2 = q$, $q_1 \leqslant q-1$ и функция $q_1(q_1+1)(1-q_1)$ принимает максимальное значение при $q_1 = q-1$, получаем отсюда, что вероятность успеха атаки не превосходит $\varepsilon q(q-1)/2$, что и требуется доказать. ■

Рассмотрим AE-систему GCM', которая отличается от GCM лишь тем, что на шаге вычисления метки сумма $GHASH(H, A, C) \oplus E_K(Y_0)$ заменяется результатом зашифрования $E_K(GHASH(H, A, C))$. Это соответствует замене системы $WC[G, F]$ системой $FH[G, F]$.

Как и в теореме 1, ограничимся случаем, когда вектор iv имеет фиксированную длину. В этих условиях имеет место следующий аналог теоремы 1.

Теорема 2. Пусть $\mathcal{C}$ — противник, имеющий преимущество $\mathcal{C}_{\text{GCM'}}$ в атаке различения семейства функций, реализуемого GCM', или в активной атаке против GCM', при числе запросов к оракулам, не превосходящем q. Пусть для каждого запроса (iv, A, P) выполняются условия $l(A) + l(C) \leqslant l$ и $l(iv) = n - t$. Тогда существует различитель $\mathcal{B}$ базового шифра E, имеющий преимущество $\mathcal{B}_E$, где

$$\mathcal{C}_{\text{GCM'}} \leqslant \mathcal{B}_E + \frac{q(q+1)}{2^{n+1}}\left\{\left\lceil \frac{l}{n} + 1\right\rceil 2^{n-\tau} + 1\right\}.$$

Доказательство. Воспользуемся схемой доказательства теоремы 1. Получим сначала оценку стойкости аутентификации. Пусть $\mathcal{C}$ — противник, имеющий преимущество $\mathcal{C}_{GCM'}$ в активной атаке против GCM', и $\mathcal{B}$ — различитель семейства функций E, построенный так же, как в доказательстве теоремы 1. Преимущество $\mathcal{B}_{\text{PRF}}$ определяется формулой (3). Из описания действий различителя $\mathcal{B}$ следует, что $\mathcal{C}_{\text{GCM'}}$ выражается формулой (4).

При получении оценки вероятности $\mathsf{P}[d^{\mathcal{B}} = 1|W_0]$ заметим, что при условии W_0 реализуется система MAC $FH[G]$. Согласно лемме 1, $G = GHASH$ образует εXU-семейство функций (следовательно, и εU-семейство), где $\varepsilon = \lceil l/n+1\rceil 2^{-\tau}$; $l(A)+l(C) \leqslant l$.

Из утверждения 4 получаем неравенство

$$\mathsf{P}\left[d^{\mathcal{B}} = 1|W_0\right] \leqslant \left\lceil \frac{l}{n} + 1\right\rceil \frac{q(q-1)}{2^{\tau+1}}.$$

Теперь из (8) и (9), с учётом леммы 2, получаем соотношения

$$\mathcal{C}_{\text{GCM'}} = \mathcal{B}_{\text{PRF}} + \mathsf{P}\left[d^{\mathcal{B}} = 1|W_0\right] \leqslant \mathcal{B}_{\text{PRF}} + \frac{\lceil l/n+1\rceil\, q(q-1)}{2^{\tau+1}} \leqslant$$
$$\leqslant \mathcal{B}_E + \frac{\lceil l/n+1\rceil\, q(q-1)}{2^{\tau+1}} + \frac{q(q-1)}{2^{n+1}},$$

откуда следует искомое неравенство.

Оценка стойкости шифрования GCM' получается так же, как в теореме 1, с той лишь разницей, что в (12) ссылка на утверждение 2 заменяется ссылкой на утверждение 4. ■

Проиллюстрируем полученную оценку. Пусть длина каждого сообщения не превосходит 12000 битов, $l(iv) = 96$, $\tau = 96$, $n = 128$. Тогда преимущество атакующего GCM' не превосходит $1{,}1 \cdot 10^{-8}$, если базовый блочный шифр неотличим от истинно случайной подстановки при защите не более чем 2^{32} сообщений на одном ключе (т. е. если при этом $\mathcal{B}_E$ практически равно нулю).

Отметим, что аналогичную оценку стойкости криптосистемы GCM' в общем случае (когда вектор iv имеет произвольную длину) можно получить, используя подход, предложенный в [12].

Как видим, оценку стойкости GCM' можно считать удовлетворительной, хотя она уступает оценке стойкости GCM. Вместе с тем GCM' защищена от атаки Фергюсона [13], которая, по мнению ряда специалистов, представляет собой основную известную угрозу для GCM. Это делает криптосистему GCM' интересной для дальнейшего изучения.

ЛИТЕРАТУРА

1. *Black J.* Message authentication codes. PhD Dissertation. Dept. of Comp. Sciences, US Davis, 2000. 126 p. `http://www.cs.colorado.edu/~jrblack/`
2. *Stinson D.* Universal hash families and the leftover hash lemma, and applications to cryptography and computing // J. Combin. Math. Combin. Comput. 2001. V. 42. No. 3. 29 p.
3. *Bellare M. and Namprempre C.* Authenticated encryption: relations among notions and analysis of the composition paradigm // Asiacrypt 2000. LNCS. 2000. V. 1976. P. 541–545.
4. CAESAR: competition for authenticated encryption: security, applicability, and robustness. 2012. `http://competitions.cr.yp.to/caesar.html`
5. *Chakraborty D. and Sarkar P.* On modes of operations of a block cipher for authentication and authenticated encryption. Cryptology ePrint Archive: report 627/14. 2014. 51 p.
6. *Rogaway P.* Authenticated-encryption with associated-data. ACM CCS, ACM Press, 2002. 10 p.
7. *Svenda P.* Basic Comparison of Modes for Authenticated-Encryption (IAPM, XCBC, OCB, CCM, EAX, CWC, GCM, PCFB, CS). 2005. 16 p. `https://www.fi.muni.cz/~xsvenda/docs/AE_comparison_ipics04.pdf`
8. *McGrew D. A. and Viega J.* The security and performance of Galois/Counter mode of operation // LNCS. 2004. V. 3348. P. 343–355.
9. *Bellare M.* Practice-oriented provable-security // LNCS. 2003. V. 1561. P. 1–15.
10. *Shrimpton T.* A characterization of authenticated-encryption as a form of chosen-ciphertext security. Cryptology ePrint Archive: 2004/272. 2004. 7 p.
11. *Bellare M., Kilian J., and Rogaway P.* The security of the cipher block chaining // LNCS. 1994. V. 839. P. 341–358.
12. *Iwata T., Ohashi K., and Minematsu K.* Breaking and repairing GCM security proofs // Crypto 2012. LNCS. 2012. V. 7417. P. 31–49.
13. *Ferguson N.* Authentication weaknesses in GCM. Public Comments to NIST. `http://csrc.nist.gov/CryptoToolKit/modes/comments`, May 2005.

REFERENCES

1. *Black J.* Message authentication codes. PhD Dissertation. Dept. of Comp. Sciences, US Davis, 2000. 126 p. `http://www.cs.colorado.edu/~jrblack/`

2. *Stinson D.* Universal hash families and the leftover hash lemma, and applications to cryptography and computing. J. Combin. Math. Combin. Comput., 2001, vol. 42, no. 3. 29 p.
3. *Bellare M. and Namprempre C.* Authenticated encryption: relations among notions and analysis of the composition paradigm. Asiacrypt 2000, LNCS, 2000, vol. 1976, pp. 541–545.
4. CAESAR: competition for authenticated encryption: security, applicability, and robustness. 2012. `http://competitions.cr.yp.to/caesar.html`
5. *Chakraborty D. and Sarkar P.* On modes of operations of a block cipher for authentication and authenticated encryption. Cryptology ePrint Archive: report 627/14. 2014. 51 p.
6. *Rogaway P.* Authenticated-encryption with associated-data. ACM CCS, ACM Press, 2002. 10 p.
7. *Svenda P.* Basic Comparison of Modes for Authenticated-Encryption (IAPM, XCBC, OCB, CCM, EAX, CWC, GCM, PCFB, CS). 2005. 16 p. `https://www.fi.muni.cz/~xsvenda/docs/AE_comparison_ipics04.pdf`
8. *McGrew D. A. and Viega J.* The security and performance of Galois/Counter mode of operation. LNCS, 2004, vol. 3348, pp. 343–355.
9. *Bellare M.* Practice-oriented provable-security. LNCS, 2003, vol. 1561, pp. 1–15.
10. *Shrimpton T.* A characterization of authenticated-encryption as a form of chosen-ciphertext security. Cryptology ePrint Archive: 2004/272. 2004. 7 p.
11. *Bellare M., Kilian J., and Rogaway P.* The security of the cipher block chaining. LNCS, 1994, vol. 839, pp. 341–358.
12. *Iwata T., Ohashi K., and Minematsu K.* Breaking and repairing GCM security proofs. Crypto 2012, LNCS, 2012, vol. 7417, pp. 31–49.
13. *Ferguson N.* Authentication weaknesses in GCM. Public Comments to NIST. `http://csrc.nist.gov/CryptoToolKit/modes/comments`, May 2005.

УДК 512.772.7

ГРАНИЦЫ СБАЛАНСИРОВАННОЙ СТЕПЕНИ ВЛОЖЕНИЯ ДЛЯ КРИПТОГРАФИИ НА БИЛИНЕЙНЫХ СПАРИВАНИЯХ

С. А. Новоселов

Балтийский федеральный университет им. И. Канта, г. Калининград, Россия

Вводится формула для расчёта границ сбалансированной степени вложения гиперэллиптической кривой. Вычислены текущие границы для кривых рода 1–3. Для кривых с известными алгоритмами генерации, наименьшими ρ-значениями и степенями вложения от 1 до 10 вычислен диапазон значений, которому принадлежит уровень безопасности кривой.

Ключевые слова: *криптография, гиперэллиптические кривые, задача дискретного логарифмирования, билинейные спаривания, степень вложения.*

DOI 10.17223/20710410/32/5

ON BOUNDS FOR BALANCED EMBEDDING DEGREE

S. A. Novoselov

Immanuel Kant Baltic Federal University, Kaliningrad, Russia

E-mail: snovoselov@kantiana.ru

A generalized formula for calculating bounds for the balanced value of the hyperelliptic curve embedding degree is proved. Using this formula we give bounds for curves of genus 1–3 over finite fields with the small, medium and big characteristic. We also compute possible range of security level for curves with known generation methods, minimal ρ-value and embedding degrees $k = 1, 2, \ldots, 10$.

Keywords: *hyperelliptic curve cryptography, pairings, embedding degree, discrete logarithm problem.*

Введение

Криптография, основанная на билинейных спариваниях, — активно развивающаяся в настоящее время область исследований. Впервые билинейные спаривания были использованы в криптографии А. Менезесом, Т. Окамото и С. А. Ванстоуном для проведения атаки на суперсингулярные эллиптические кривые путём сведения дискретного логарифма на эллиптической кривой к дискретному логарифму в конечном поле [1], в котором существуют более эффективные алгоритмы вычисления дискретного логарифма. Данный метод подходит также и для других классов кривых, на которых возможно эффективное вычисление спариваний.

Позже в 2000 г. были предложены первые конструктивные приложения спариваний — схема цифровой подписи и схема распределения ключей [2], трёхсторонний протокол Диффи — Хеллмана [3]. После этого было разработано большое число криптосистем, основанных на спариваниях, среди которых выделяется IBE [4]. Обзор таких криптосистем можно найти в работе [5].

Безопасность и эффективность данных криптосистем зависит от двух параметров — степени вложения k и минимальной степени вложения k'. Степень вложения k определяет размер поля, над которым вычисляется билинейное спаривание; соответственно от данного параметра зависит скорость работы криптосистемы. Минимальная степень вложения k', введённая в работе [6], определяет минимальный размер поля, к которому сводится задача вычисления дискретного логарифма в якобиане кривой и, следовательно, является параметром, размер которого определяет безопасность криптосистемы. Эти параметры совпадают в случае, если кривая задана над простым конечным полем. Для случая непростого поля некоторые условия даны в работе [7]. Если параметры совпадают, то k является характеристикой безопасности кривой.

Степень вложения должна быть достаточно большой, чтобы сведение к конечному полю не позволяло решить проблему за меньшее время. С другой стороны, размер степени вложения влияет на эффективность вычисления функции спаривания, которая является основой для криптосистем на спариваниях, поэтому для построения эффективных криптосистем степень вложения должна быть как можно меньше. В связи с этим необходимо выбирать сбалансированное значение k — такое, что сложность решения задачи вычисления дискретного логарифма в якобиане гиперэллиптической кривой равна сложности решения задачи в конечном поле.

Для эллиптических кривых в работе [8] есть асимптотическая оценка, рассчитанная для актуальных по состоянию на 2008 г. алгоритмов дискретного логарифмирования в конечном поле со сложностью $\approx e^{(\ln q)^{1/3}}$. В связи с появлением принципиально новых методов дискретного логарифмирования в конечном поле [9, 10] эта оценка больше не является верной в общем случае.

Целью данной работы является вывод более общей формулы для оценки размера сбалансированного значения k для гиперэллиптических кривых, включающей в себя оценки других авторов как частный случай. Кроме того, проводится сравнение размера сбалансированного значения для разных кривых, различных комбинаций используемых алгоритмов решения задачи вычисления дискретного логарифма и разных типов конечных полей.

Работа организована следующим образом. П. 1 содержит основные определения, предварительные сведения и оценки других авторов для сбалансированной степени вложения. В п. 2 проводится обзор алгоритмов решения задачи вычисления дискретного логарифма в исследуемых группах с выбором наилучших на момент написания работы. Оценки сложности данных алгоритмов и значения констант в них используются для расчётов и вывода формул в последующих пунктах. В п. 3 выводятся общие формулы для расчёта границ сбалансированной степени вложения, позволяющие получить диапазон значений, которому она принадлежит. В п. 4 вводится понятие уровня безопасности, которое необходимо для сравнения кривых. Выводятся формулы для расчёта сбалансированной степени вложения, размеров групп и полей, необходимых для обеспечения заданного уровня безопасности. В п. 5 на основе предыдущих результатов выводится формула для вычисления диапазона уровней безопасности, которому принадлежит уровень безопасности заданной эллиптической или гиперэллиптической кривой с некоторой фиксированной степенью вложения. Кроме того, по выведенной формуле рассчитываются диапазоны уровней безопасности для кривых, предложенных в литературе для использования в криптографии на билинейных спариваниях, и выделяются классы кривых, небезопасные для использования в настоящее время.

1. Предварительные сведения

Определение 1. Пусть G_1, G_2, G_3 — группы порядка n. Билинейным спариванием называется отображение

$$e_n : G_1 \times G_2 \to G_3$$

со следующими свойствами:

1) билинейность:

$$\forall P \in G_1 \, \forall Q, R \in G_2 \ (e_n(P, Q + R) = e_n(P, Q)e_n(P, R)),$$
$$\forall P, R \in G_1 \, \forall Q \in G_2 \ (e_n(P + R, Q) = e_n(P, Q)e_n(R, Q));$$

2) невырожденность:

$$\forall P \in G_1 \setminus \{0\} \ \exists Q \in G_2 \ (e_n(P, Q) \neq 1),$$
$$\forall Q \in G_2 \setminus \{0\} \ \exists P \in G_1 \ (e_n(P, Q) \neq 1).$$

Пусть C — гиперэллиптическая кривая рода g; r — простое число, такое, что $r|\#Jac_C(\mathbb{F}_q)$.

Большинство билинейных спариваний являются модификациями спаривания Тейта — Лихтенбаума, которое определяется как невырожденное билинейное отображение:

$$Jac_C(\mathbb{F}_{q^k})[r] \times Jac_C(\mathbb{F}_{q^k})/rJac_C(\mathbb{F}_{q^k}) \to \mathbb{F}_{q^k}^{\times}/(\mathbb{F}_{q^k}^{\times})^r.$$

Число k называется степенью вложения кривой C. Если r и q взаимно просты, то степень вложения определяется как наименьшее целое k, такое, что $r|q^k - 1$. Если q не взаимно просто с r, то имеем $r = p$, и подгруппа порядка r не подходит для криптографии на спариваниях, так как задача вычисления дискретного логарифма в $Jac_C(\mathbb{F}_{q^k})[p]$ решается за полиномиальное время от $\log q$ [11].

В [6] введён второй параметр безопасности k', который называется минимальной степенью вложения. Если $q = p^m$, то $k' = ord_r p/m \in \mathbb{Q}$. Этот параметр определяет минимальное поле вложения $\mathbb{F}_{q^{k'}}$, в котором содержится группа корней из единицы.

На практике спаривание вычисляется с помощью алгоритма Миллера [12] и его модификаций, которые имеют сложность $\mathrm{O}(\log r)$. Так как спаривание вычисляется над полем $\mathbb{F}_{q^k}$, сложность операции в котором $\mathrm{O}(k \log q)$, для эффективности вычислений k должно быть мало — асимптотически как полиномиальная функция от битового размера поля: $k = O((\log q)^{c_k})$.

Известно, что большая часть кривых имеет экспоненциальную степень вложения. Небольшую степень вложения имеют малые классы кривых, многие из которых специально подобраны для использования в криптосистемах на спариваниях. Обзор и методы построения таких кривых есть в работах [13, 14] и п. 5.5.

Так как с помощью спариваний дискретный логарифм в группе точек эллиптической кривой или в якобиане гиперэллиптической кривой сводится к дискретному логарифму в конечном поле, для безопасного использования криптосистем на спариваниях должно выполнятся неравенство

$$C_{DLP}(\mathbb{F}_{q^k}^{\times}) \geqslant C_{DLP}(Jac_C(F_q)),$$

где $C_{DLP}(G)$ — сложность решения дискретного логарифма в группе G. Вследствие того, что при увеличении k увеличивается время вычисления спаривания, для приложений необходимо, чтобы k было сбалансировано, т. е. выполнялось равенство

$$C_{DLP}(F_{q^k}^{\times}) \approx C_{DLP}(Jac_C(F_q)).$$

В случае эллиптических кривых в [8] есть оценка

$$k \approx \alpha(c,\rho)\left(\frac{\log r}{\log\log r}\right)^2, \tag{1}$$

где $\rho = \log q/\log r$; $\alpha \approx 1/100\rho$. Заметим, что эта оценка рассчитывалась для алгоритмов дискретного логарифмирования в конечном поле со сложностью $L_q(1/3, c)$.

В 2013 г. появилось улучшение метода исчисления индексов [9], на порядок уменьшающее сложность в случае малой характеристики с $L_q(1/3)$ до $L_q(1/4)$. Кроме того, появился новый теоретический квазиполиномиальный алгоритм [10]. Поэтому оценка k из формулы (1) больше не работает в общем случае.

2. Задача вычисления дискретного логарифма

2.1. Общий случай

Пусть G — аддитивная группа порядка n; g, h — элементы этой группы. Задача вычисления дискретного логарифма (ВДЛ) состоит в следующем: по паре (g, h) найти такое число l, что $h = lg$, если такое число существует. В случае мультипликативной группы задача формулируется аналогично.

В общем случае задача ВДЛ решается за время $\mathrm{O}(\sqrt{n})$ с помощью алгоритма Шенкса или ρ-метода Полларда [15]. Если $n = \prod\limits_i p_i^{e_i}$ — разложение порядка группы на простые множители, то задача решается за время $\mathrm{O}\left(\sum\limits_i e_i(\log n + \sqrt{p_i})\right) = \mathrm{O}(\max\sqrt{p_i})$ с помощью алгоритма Полига — Хеллмана [16]. Поэтому далее ограничимся случаем, когда порядок группы n — большое простое число, которое обозначим r; соответственно в данном случае сложность ВДЛ равна $\mathrm{O}(\sqrt{r})$.

2.2. Конечные поля

Пусть $\mathbb{F}_q$ — конечное поле; $q = p^m$; p — простое число.

В конечных полях существуют более эффективные алгоритмы решения задачи ВДЛ, основанные на методе исчисления индексов — метод решета числового поля (NFS), метод решета функционального поля (FFS) и их модификации. Сложность решения и выбор оптимального алгоритма зависят от соотношения между характеристикой p и степенью поля m. Для выражения сложности этих алгоритмов, а также соотношения между p и m используется L-нотация:

$$L_x(\alpha, c) = e^{(\ln x)^\alpha (\ln\ln x)^{1-\alpha}(c+\mathrm{o}(1))},$$

где $0 \leqslant \alpha \leqslant 1$; $c > 0$. При этом:

1) если $\alpha = 0$, то сложность полиномиальная от $\ln q$;
2) если $\alpha = 1$, то сложность экспоненциальная;
3) если $0 < \alpha < 1$, то сложность субэкспоненциальная.

Параметр α наиболее важен, так как он определяет состояние между экспоненциальной сложностью и полиномиальной. Поэтому часто используется сокращенный вариант записи $L_x(\alpha)$, при котором константа c опускается.

Обозначим соотношение между p и m как $p = L_{p^m}(l_p, c)$. Выделяют [17] три случая:

1) малая характеристика: $l_p \leqslant 1/3$;
2) средняя характеристика: $1/3 \leqslant l_p \leqslant 2/3$;
3) большая характеристика: $l_p \geqslant 2/3$.

В граничных случаях доступно несколько алгоритмов, и выбирается наилучший.

Малая характеристика

В случае малой характеристики наилучший алгоритм имеет сложность $L_q(1/4 + o(1), c)$ [9]. Существует также теоретический алгоритм [10], имеющий квазиполиномиальное время $2^{O((\log\log q)^2)}$, который, однако, имеет большой размер константы. Данные алгоритмы основаны на недоказанных формально гипотезах, но получили подтверждение при вычислениях дискретных логарифмов на практике [18, 19]. Поэтому в настоящее время поля малой характеристики считаются небезопасными для криптографии, основанной на дискретном логарифмировании. С историей вопроса можно ознакомиться в работе [20].

Средняя и большая характеристика

В случае средней характеристики наилучшие алгоритмы имеют сложность $L_q(1/3)$. Наиболее эффективные алгоритмы — MNFS [21] с константой $c = \left(8(9+4\sqrt{6})/15\right)^{1/3} \approx$ $\approx 2{,}156$ и exTNFS [22] с константой $c = \sqrt[3]{48/9} = 1{,}747$.

Для большой характеристики наилучший алгоритм — также MNFS [23] со сложностью в этом случае

$$L_q\left(\frac{1}{3}, \left(\frac{92+26\sqrt{13}}{27}\right)^{1/3}\right).$$

Информация по алгоритмам решения задачи ВДЛ в конечных полях приведена в табл. 1.

Таблица 1

Алгоритмы дискретного логарифмирования в конечных полях

Алгоритм	Характеристика	Сложность
BGJT [10]	Малая	$2^{O((\ln\ln q)^2)}$
Joux [9]	Малая	$L_q(1/4 + o(1), c)$
MNFS [21]	Средняя	$L_q(1/3, 2{,}156)$
exTNFS [22]	Средняя	$L_q(1/3, 1{,}747)$
MNFS [23]	Большая	$L_q(1/3, 1{,}901)$

2.3. Эллиптические кривые

Для эллиптических кривых в общем случае лучший алгоритм — ρ-метод Полларда, имеющий сложность $O(\sqrt{r})$. В специальных случаях возможно сведение проблемы к конечному полю [1, 24], изучаемое в данной работе, либо к кривым высокого рода с использованием спуска Вейля [25, 11]. Последний метод применим только для композитных полей с числом элементов $q = p^m$, где m — составное число. При этом в общем случае методы, основанные на спуске Вейля, не обязательно ведут к более эффективному решению задачи, но для использования кривых над композитными полями в криптографии необходимо доказывать неприменимость данной атаки. Поэтому обычно требуется, чтобы m было простым числом. Заметим, что спуск Вейля также применим и к более общему случаю гиперэллиптических кривых. Для некоторых классов эллиптических кривых доступен метод исчисления индексов [26] с асимптотической сложностью

$$2^{c\sqrt{\log q \log\log q}} = L_q(1/2, c\ln 2), \quad c \approx 1{,}69.$$

2.4. Гиперэллиптические кривые

Пусть C — гиперэллиптическая кривая рода g над полем $\mathbb{F}_q$, $G \subseteq Jac_C(\mathbb{F}_q)$ — подгруппа простого порядка r.

В общем случае задача дискретного логарифмирования на гиперэллиптических кривых может быть решена с помощью ρ-метода Полларда, имеющего сложность $\mathrm{O}(\sqrt{r})$. Однако для кривых рода $g \geqslant 3$ существуют более эффективные методы, основанные на методе исчисления индексов.

Кривые рода 2

Для кривых рода 2 в общем случае наилучший алгоритм — ρ-метод Полларда со сложностью $\mathrm{O}(\sqrt{r}) = O(q)$ (табл. 2). Поэтому гиперэллиптические кривые рода 2 и эллиптические кривые в настоящее время считаются наиболее подходящими для криптографии.

Таблица 2

Алгоритмы дискретного логарифмирования на гиперэллиптических кривых

Алгоритм	Род g	Сложность
ρ-метод Полларда	Любой	$L_r(1, 0{,}5)$
Semaev [26]	1	$L_r(1/2, 1{,}171)$
GTTD [27]	$g \geqslant 3$, $q > g!$	$L_r(1, (2-2/g)\rho/g)$
Enge — Gaudry [28]	$g > \mathrm{O}(\log q)$	$L_r(1/2, 1{,}414)$
Enge — Gaudry — Thome [29]	$g > \mathrm{O}(\log q)^2$	$L_r(1/3, 1{,}922)$

Кривые малого рода $g \geqslant 3$

В случае малого рода $g \geqslant 3$ существуют методы ВДЛ со сложностью $\tilde{\mathrm{O}}(q^{2-2/g})$, если $q > g!$ [27]. Заметим, что данный алгоритм зависит от ρ-значения — соотношения между размером якобиана и размером подгруппы — и превосходит ρ-метод Полларда только при $\rho < g^2/(4(g-1))$.

Кривые большого рода

Если род g достаточно большой по сравнению с q, то задача решается за субэкспоненциальное время от q^g с помощью метода исчисления индексов.

Теорема 1 (Enge — Gaudry [28]). Предположим, что $g > \vartheta \log(q)$, где ϑ — константа. Тогда существует алгоритм, который решает задачу ВДЛ на гиперэллиптических кривых за время $\mathrm{O}(L_{q^g}(1/2, c(\vartheta)))$, где q, g, ϑ стремятся к бесконечности, а $c(\vartheta)$ — функция, стремящаяся к $\sqrt{2}$.

3. Сбалансированное значение k

Пусть C — гиперэллиптическая кривая рода g над полем $\mathbb{F}_q$, $G \subseteq Jac_C(\mathbb{F}_q)$ — подгруппа простого порядка r.

Теорема 2 (Хассе — Вейль). Имеет место неравенство

$$\left(q^{1/2} - 1\right)^{2g} \leqslant |Jac_C(\mathbb{F}_q)| \leqslant \left(q^{1/2} + 1\right)^{2g}.$$

Поэтому $r \approx q^g$ при $q \to \infty$. Обозначим и зафиксируем $\rho = g \log q / \log r$. Это значение измеряет отношение между размером подгруппы, выбранной для криптосистемы, и размером якобиана кривой. Заметим, что $\rho \geqslant 1$. Для построения криптосистем на

практике это значение должно быть мало, чтобы уменьшить расходы на операции в группе.

Если $(r, q) = 1$, то степень вложения определяется как минимальное k, такое, что $r|q^k - 1$. Поэтому выполняется

Утверждение 1. Пусть C — гиперэллиптическая кривая рода g над полем $\mathbb{F}_q$; G — подгруппа Якобиана кривой простого порядка r; $(r, q) = 1$; k — степень вложения; $\rho = g \log q / \log r$. Тогда выполняется следующее неравенство:

$$k > g/\rho.$$

Доказательство. Так как $r|q^k - 1$, то $q^k - 1 \geqslant r$. Значит, $q^k > r$, $k > \log r / \log q = g/\rho$. ■

Аналогично можно доказать утверждение для минимальной степени вложения: $k' > g/\rho$.

Обозначим $L_{q^k}(\alpha, c_1)$ и $L_r(\beta, c_2)$ — сложности решения задачи дискретного логарифмирования в конечном поле и в подгруппе якобиана кривой соответственно.

В случае эллиптических кривых в [12, 30] указывается $k < \ln^2 q$ как необходимое условие для субэкспоненциальности (от $\ln q$) проблемы ВДЛ в конечном поле $\mathbb{F}_{q^k}$. Это условие верно для алгоритмов с $\alpha = 1/3$. Обобщая условие на случай произвольного $\alpha \in (0, 1)$, получаем

$$k < (\ln q)^{1/\alpha - 1}.$$

В случае гиперэллиптических кривых необходимо учитывать, что $r \approx q^{g/\rho}$. Заметим, что при $k = (g/\rho)^{1/\alpha} (\ln q)^{1/\alpha - 1} = (g/\rho)(\ln r)^{1/\alpha - 1}$ имеем

$$L_{q^k}(\alpha, c_1) = L_r(1, \alpha^{\alpha - 1} c_1)^{(\ln \ln r)^{1-\alpha}},$$

что больше $L_r(\beta, c_2)$ для любого β. Соответственно сбалансированное значение k ограничено следующим образом:

$$\frac{g}{\rho} < k < \frac{g}{\rho} (\ln r)^{1/\alpha - 1}. \tag{2}$$

Следовательно, $k = (g/\rho)(\ln r)^{c_k}$ для некоторого $0 < c_k < 1/\alpha - 1$. Таким образом, чтобы гарантировать стойкость кривой к MOV/FR-атаке, достаточно выбрать кривую со степенью вложения $k \geqslant (g/\rho)(\ln r)^{1/\alpha - 1}$. Для большинства кривых это условие выполняется, так как k в общем случае имеет размер, близкий к r.

Заметим, что на практике в настоящее время $\alpha = 1/3$ или $1/4$ для любых конечных полей [17]. Вследствие этого получаем гарантии безопасности криптосистемы при $k \geqslant (g/\rho)(\ln r)^2$ или $k \geqslant (g/\rho)(\ln r)^3$ соответственно. Однако кривые с такой степенью вложения не подходят для криптосистем на спариваниях, так как значения k слишком большие.

Для использования кривой в криптосистемах на спариваниях необходимо, чтобы сложности решения задачи вычисления дискретного логарифма в якобиане кривой и в конечном поле были сбалансированы:

$$L_{q^k}(\alpha, c_1) \approx L_r(\beta, c_2). \tag{3}$$

Заметим, что по свойствам L-нотации [31] для любых констант $c > 0$, a и δ выполняется $(\ln x)^a L_x(\delta, c) = L_x(\delta, c)$ и $L_x(\delta, c) L_x(\gamma, b) = L_x(\delta, c)$, если $\delta > \gamma$. Это связано

с тем, что $L_x(\delta, c)$ — сокращение для $L_x(\delta, c + \mathrm{o}(1))$ и множители $(\ln x)^a$ и $L_x(\gamma, b)$ при соответствующем преобразовании выражения попадают в $\mathrm{o}(1)$. При этом исходные константы c и δ не изменяются, меняется только константа в $\mathrm{o}(1)$.

Поэтому в оценках сложности $L_{q^k}(\alpha, c_1)$ и $L_r(\beta, c_2)$ решения задачи вычисления дискретного логарифма все временные затраты на вспомогательные операции, такие как операции в конечном поле, сложность сведения задачи к конечному полю и другие, попадают в $\mathrm{o}(1)$ и не влияют на константы c_1, c_2, α, β, если совокупное время их выполнения при работе алгоритма асимптотически не больше чем $L_{q^k}(\delta, c)$, $\delta < \alpha$, в первом случае и $L_r(\delta, c)$, $\delta < \beta$, — во втором. Если эти условия выполняются, то можно считать, что сложности $L_{q^k}(\alpha, c_1)$ и $L_r(\beta, c_2)$ выражены в битовых операциях, а не в операциях в поле $\mathbb{F}_{q^k}$ и якобиане кривой.

Покажем, что эти условия выполняются. Заметим, что групповая операция в якобиане кривой в общем случае вычисляется с помощью алгоритма Кантора [32] и его оптимизированных версий либо с помощью более быстрых явных формул при их наличии (например, если $g = 1$ или 2).

Утверждение 2 [33, § 2.6]. Если поле $\mathbb{F}_q$ имеет нечётную характеристику, то групповая операция в $Jac_C(\mathbb{F}_q)$ может быть вычислена за $17g^2 + \mathrm{O}(g)$ операций в поле $\mathbb{F}_q$. В случае чётной характеристики групповая операция может быть вычислена за $14g^2 + \mathrm{O}(g)$ операций.

Таким образом, операция в якобиане кривой имеет сложность $\mathrm{O}(g^2)$ операций в поле $\mathbb{F}_q$, или $\mathrm{O}(g^2(\log q)^2)$ битовых операций. Необходимо также учитывать затраты на выполнение сведения задачи ВДЛ к конечному полю. Общий алгоритм сведения для эллиптических кривых можно найти в [34, IX.9], для гиперэллиптических кривых — в [11].

Утверждение 3. Пусть C — гиперэллиптическая кривая рода g над полем $\mathbb{F}_q$; $G \subseteq Jac_C(\mathbb{F}_q)$; $|G| = r$ — простое число; k — степень вложения. Пусть $L_{q^k}(\alpha, c_1)$ — сложность решения задачи вычисления дискретного логарифма в поле $\mathbb{F}_{q^k}$, выраженная в операциях в этом поле. Тогда сведение задачи дискретного логарифмирования из якобиана кривой в конечное поле $\mathbb{F}_{q^k}$ имеет сложность $L_{q^k}(\alpha, c_1)$ битовых операций.

Доказательство. Покажем, что при выражении сложности в битовых операциях вспомогательные операции не влияют на константы α, c в выражении для сложности алгоритма вычисления дискретного логарифма в конечном поле. Кроме того, на эти константы также не оказывает влияния род кривой g.

Для сведения задачи ВДЛ к конечному полю необходимо вычислять функцию спаривания над полем $\mathbb{F}_{q^k}$, затем вычислять дискретный логарифм в этом поле за время $L_{q^k}(\alpha, c_1)$. При этом процесс может завершиться неудачей, если результатом вычисления спаривания будет элемент меньшего порядка, чем нужно. В этом случае выбирается другая случайная вспомогательная точка, и процесс повторяется. Ожидаемое число попыток для успешного применения атаки — $\mathrm{O}(\ln \ln r)$ [1]. Сложность операций в конечном поле $\mathbb{F}_{q^k}$ в общем случае — $\mathrm{O}((\ln q^k)^2)$ битовых операций.

Сложность вычисления спаривания — $\mathrm{O}(\ln r)$ операций в якобиане кривой $Jac_C(\mathbb{F}_q)$ [12; 24, Prop. 3.2]. Применяя утверждение 2, получаем сложность вычисления спаривания в битовых операциях $\mathrm{O}(g^2 \ln r (\ln q^k)^2)$. В итоге сложность атаки в битовых операциях составляет

$$(\mathrm{O}(g^2 \ln r) + L_{q^k}(\alpha, c_1))\mathrm{O}((\ln q^k)^2 \ln \ln r). \tag{4}$$

Заметим, что по теореме 2 имеем $r \leqslant (\sqrt{q} + 1)^{2g}$, то есть $r = \mathrm{O}(q^g)$. Кроме того, так как $k > g/\rho$, то $g = \mathrm{O}(k)$. Следовательно, $\ln r = \mathrm{O}(\ln q^k)$ и $\ln \ln r = \mathrm{O}(\ln \ln q^k)$. Поэтому

выражение (4) можно записать в виде

$$(\mathrm{O}(k^2 \ln q^k) + L_{q^k}(\alpha, c_1))\mathrm{O}((\ln q^k)^2 \ln\ln q^k). \tag{5}$$

Имеем

$$\mathrm{O}(k^2 \ln q^k) = \exp(\ln C + \ln k^2 + \ln\ln q^k) = \exp\left(\ln\ln q^k \left(\frac{2\ln k}{\ln\ln q + \ln k} + 1 + \mathrm{o}(1)\right)\right) =$$

$$= \exp\left(\ln\ln q^k \left(\frac{2}{\frac{\ln\ln q}{\ln k} + 1} + 1 + \mathrm{o}(1)\right)\right) = L_{q^k}(0, c)$$

для некоторой константы $c < 3$. Подставляя это выражение в (5), получаем

$$(L_{q^k}(0,c) + L_{q^k}(\alpha, c_1))\mathrm{O}((\ln q^k)^2 \ln\ln q^k) = L_{q^k}(\alpha, c_1)\mathrm{O}((\ln q^k)^2 \ln\ln q^k) = L_{q^k}(\alpha, c_1).$$

Утверждение доказано. ■

Аналогично можно записать в битовых операциях сложность $L_r(\beta, c_2)$, выраженную в операциях в якобиане кривой над полем $\mathbb{F}_q$:

$$L_r(\beta, c_2)\mathrm{O}(g^2 \ln^2 q) = L_r(\beta, c_2)\mathrm{O}((\rho \ln r)^2) = L_r(\beta, c_2).$$

Таким образом, условие (3) можно записать не в приближённом, а в точном виде:

$$L_{q^k}(\alpha, c_1) = L_r(\beta, c_2),$$

предполагая, что сложности выражены в битовых операциях, а все вспомогательные операции попадают в o(1).

Введём теперь более точную границу для сбалансированного значения k.

Теорема 3. Пусть C — гиперэллиптическая кривая рода g над полем $\mathbb{F}_q$; $G \subseteq Jac_C(\mathbb{F})$; $|G| = r$ и r — простое число; $L_{q^k}(\alpha, c_1)$ — сложность решения задачи ВДЛ в конечном поле; $L_r(\beta, c_2)$ — сложность решения задачи ВДЛ в группе G. Тогда если $\beta < 1$ или $\beta = 1$ и $\dfrac{c_2 + \mathrm{o}(1)}{c_1 + \mathrm{o}(1)} < \left(\dfrac{1}{\alpha} - \left(\dfrac{1}{\alpha} - 1\right)\dfrac{\ln\ln\ln r}{\ln\ln r}\right)^{1-\alpha}$, то для сбалансированной степени вложения выполняется неравенство

$$\frac{g}{\rho} < k < \frac{g}{\rho}\left(\frac{\ln r}{\ln\ln r}\right)^{1/\alpha - 1}. \tag{6}$$

Доказательство. Подставляя $k = \dfrac{g}{\rho}\left(\dfrac{\ln r}{\ln\ln r}\right)^{1/\alpha-1}$ и $\ln q = \dfrac{\rho}{g}\ln r$ в $L_{q^k}(\alpha, c_1)$, получаем

$$e^{(c_1 + \mathrm{o}(1))\ln r\left(1/\alpha - (1/\alpha - 1)\frac{\ln\ln\ln r}{\ln\ln r}\right)^{1-\alpha}} = L_r\left(1, (c_1 + \mathrm{o}(1))\left(\frac{1}{\alpha} - \mathrm{o}(1)\right)^{1-\alpha}\right).$$

Это выражение превосходит $L_r(\beta, c_2)$, только если $\beta < 1$ или $\beta = 1$ и $\dfrac{c_2 + \mathrm{o}(1)}{c_1 + \mathrm{o}(1)} < (1/\alpha - \mathrm{o}(1))^{1-\alpha}$. ■

При $\alpha = 0$ или $\beta = 0$ кривая считается непригодной для криптографии, так как задача ВДЛ в этом случае имеет полиномиальную сложность. При $\alpha > \beta$ сложность

решения задачи в конечном поле асимптотически выше, и в этом случае для построения криптосистем лучше подходят конечные поля, так как вычисления в них проще при меньшей, в данном случае, необходимой длине ключа.

Следующая теорема позволяет вычислить асимптотические границы для величины k, при которых уровни безопасности в якобиане кривой и конечном поле сбалансированы.

Теорема 4. Пусть C — гиперэллиптическая кривая рода g и $L_{q^k}(\alpha, c_1), 0 < \alpha < 1$, $L_r(\beta, c_2), \alpha \leqslant \beta \leqslant 1$, — сложности решения задачи ВДЛ в конечном поле $\mathbb{F}_{q^k}$ и подгруппе простого порядка r якобиана кривой $Jac_C(\mathbb{F}_q)$ соответственно. Тогда уровни безопасности сбалансированы при

$$k = \frac{g}{\rho}\left(\frac{\ln r}{\ln\ln r}\right)^{\beta/\alpha-1} c,$$

где c — некоторая константа, такая, что

$$\alpha^{1/\alpha-1}\left(\frac{c_2+\mathrm{o}(1)}{c_1+\mathrm{o}(1)}\right)^{1/\alpha} < c < \left(\frac{c_2+\mathrm{o}(1)}{c_1+\mathrm{o}(1)}\right)^{1/\alpha}. \tag{7}$$

Если выполняется неравенство (6) из теоремы 3, то имеет место

$$\left(\frac{\alpha}{1-(1-\alpha)\dfrac{\ln\ln\ln r}{\ln\ln r}}\right)^{1/\alpha-1}\left(\frac{c_2+\mathrm{o}(1)}{c_1+\mathrm{o}(1)}\right)^{1/\alpha} < c < \left(\frac{c_2+\mathrm{o}(1)}{c_1+\mathrm{o}(1)}\right)^{1/\alpha}. \tag{8}$$

Доказательство. Раскрывая условие $L_{q^k}(\alpha, c_1) = L_r(\beta, c_2)$, получаем

$$e^{(\ln q^k)^\alpha(\ln\ln q^k)^{1-\alpha}(c_1+\mathrm{o}(1))} = e^{(\ln r)^\beta(\ln\ln r)^{1-\beta}(c_2+\mathrm{o}(1))}.$$

Выразим из этого выражения k:

$$(\ln q^k)^\alpha(\ln\ln q^k)^{1-\alpha}(c_1+\mathrm{o}(1)) = (\ln r)^\beta(\ln\ln r)^{1-\beta}(c_2+\mathrm{o}(1)),$$
$$k^\alpha(\ln q)^\alpha(\ln\log q+\ln k)^{1-\alpha}(c_1+\mathrm{o}(1)) = (\ln r)^\beta(\ln\ln r)^{1-\beta}(c_2+\mathrm{o}(1)),$$
$$k = \left(\frac{(\ln r)^\beta(\ln\ln r)^{1-\beta}(c_2+\mathrm{o}(1))}{(\ln q)^\alpha(\ln\ln q+\ln k)^{1-\alpha}(c_1+\mathrm{o}(1))}\right)^{1/\alpha} = \left(\frac{c_2+\mathrm{o}(1)}{c_1+\mathrm{o}(1)}\right)^{1/\alpha}\frac{(\ln r)^{\beta/\alpha}(\ln\ln r)^{(1-\beta)/\alpha}}{\ln q(\ln\ln q+\ln k)^{1/\alpha-1}}.$$

Так как $\ln q = \rho\ln r/g$, то

$$k = \left(\frac{c_2+\mathrm{o}(1)}{c_1+\mathrm{o}(1)}\right)^{1/\alpha}\frac{g}{\rho}\frac{(\ln r)^{\beta/\alpha-1}(\ln\ln r)^{(1-\beta)/\alpha}}{(\ln\ln q+\ln k)^{1/\alpha-1}}.$$

Из неравенства (2) имеем $k = \dfrac{g}{\rho}(\ln r)^{c_k}$, поэтому $\ln\ln q + \ln k = (c_k+1)\ln\ln r$. Следовательно, получаем

$$k = \left(\frac{c_2+\mathrm{o}(1)}{c_1+\mathrm{o}(1)}\right)^{1/\alpha}\left(\frac{1}{c_k+1}\right)^{1/\alpha-1}\frac{g}{\rho}\left(\frac{\ln r}{\ln\ln r}\right)^{\beta/\alpha-1}.$$

Так как $0 < c_k < 1/\alpha - 1$, получаем

$$\alpha^{1/\alpha-1} < \left(\frac{1}{c_k+1}\right)^{1/\alpha-1} < 1.$$

Если выполняется неравенство (6), то на величину c_k можно наложить дополнительные ограничения. Имеем $k = \dfrac{g}{\rho}\left(\dfrac{\ln r}{\ln\ln r}\right)^{c'_k}$ для некоторой величины c'_k, $0 < c'_k < 1/\alpha - 1$. Так как $(\ln r)^{c_k} = \dfrac{\ln r}{\ln\ln r}$, получаем

$$c_k = c'_k\left(1 - \frac{\ln\ln\ln r}{\ln\ln r}\right) < \left(\frac{1}{\alpha} - 1\right)\left(1 - \frac{\ln\ln\ln r}{\ln\ln r}\right).$$

Отсюда следует, что

$$\left(\frac{1}{c_k+1}\right)^{1/\alpha-1} > \left(\frac{1}{\left(\dfrac{1}{\alpha}-1\right)\left(1-\dfrac{\ln\ln\ln r}{\ln\ln r}\right)+1}\right)^{1/\alpha-1} = \left(\frac{\alpha}{1-(1-\alpha)\dfrac{\ln\ln\ln r}{\ln\ln r}}\right)^{1/\alpha-1}.$$

Теорема доказана. ■

Выражение $\dfrac{\ln\ln\ln r}{\ln\ln r}$ в неравенстве (8) при $r \to \infty$ стремится к нулю и соответственно всё выражение стремится к $\alpha^{1/\alpha-1}$, т. е. к выражению в неравенстве (7).

В криптографии используются значения $2^{80} < r < 2^{10240}$, что означает

$$0{,}246 < \frac{\ln\ln\ln r}{\ln\ln r} < 0{,}346.$$

Следовательно,

$$\left(\frac{\alpha}{0{,}246\alpha+0{,}753}\right)^{1/\alpha-1} < \left(\frac{\alpha}{1-(1-\alpha)\dfrac{\ln\ln\ln r}{\ln\ln r}}\right)^{1/\alpha-1} < \left(\frac{\alpha}{0{,}346\alpha+0{,}653}\right)^{1/\alpha-1}.$$

Для $\alpha = 1/3$ получаем

$$0{,}159 < \left(\frac{\alpha}{1-(1-\alpha)\dfrac{\ln\ln\ln r}{\ln\ln r}}\right)^{1/\alpha-1} < 0{,}187,$$

что больше, чем $\alpha^{1/\alpha-1} \approx 0{,}111$, и значит, граница (8) точнее.

Теорема 4 позволяет вычислить границы для сбалансированного значения k с точностью до бесконечно малых величин. Соответственно кривая со степенью вложения

$$k = \left(\frac{c_2+\mathrm{o}(1)}{c_1+\mathrm{o}(1)}\right)^{1/\alpha}\frac{g}{\rho}\left(\frac{\ln r}{\ln\ln r}\right)^{\beta/\alpha-1}$$

является стойкой к MOV/FR-атаке. В то же время если

$$k < \alpha^{1/\alpha-1}\left(\frac{c_2+\mathrm{o}(1)}{c_1+\mathrm{o}(1)}\right)^{1/\alpha}\frac{g}{\rho}\left(\frac{\ln r}{\ln\ln r}\right)^{\beta/\alpha-1},$$

то уровень безопасности на кривой меньше, чем в конечном поле, и такая кривая небезопасна для использования в криптосистемах на спариваниях. Сбалансированное значение находится среди промежуточных значений.

4. Уровень безопасности

Определение 2. Группа G обладает уровнем безопасности l, если для решения задачи ВДЛ в этой группе требуется 2^l битовых операций.

Для обеспечения одного и того же уровня безопасности в различных группах может требоваться разный размер группы; группы с меньшим требуемым размером имеют преимущество. Для выбора безопасного размера группы, удовлетворяющего заданному уровню безопасности, могут использоваться рекомендации NIST [35]. Расчёт степеней вложения на основе этих данных есть в работе [14]. В общем случае, для асимптотической оценки размера группы, необходимого для обеспечения уровня безопасности l, докажем следующую теорему.

Теорема 5. Пусть G — группа порядка n и сложность решения задачи ВДЛ в битовых операциях выражается в виде $L_n(\alpha, c)$, где $0 \leqslant \alpha \leqslant 1$; $c > 0$. Тогда размер группы $\lg n$, необходимый для обеспечения уровня безопасности l, равен

$$s_l(\alpha, c) = \begin{cases} e^{l \ln 2/(c+\mathrm{o}(1))-\ln\ln 2}, & \alpha = 0, \\ l^{1/\alpha}(\ln l)^{-(1-\alpha)/\alpha}\dfrac{(\ln 2)^{1/\alpha-1}}{(c+\mathrm{o}(1))^{1/\alpha}}\left(\dfrac{\alpha}{1-\mathrm{o}(1)}\right)^{(1-\alpha)/\alpha}, & 0 < \alpha < 1, \\ \dfrac{l}{c+\mathrm{o}(1)}, & \alpha = 1. \end{cases}$$

Доказательство.

1) $\ln^c n = 2^l$, $\ln n = e^{l\,\ln 2/(c+\mathrm{o}(1))}$.

2) Последовательно запишем:

$$(\ln n)^\alpha (\ln\ln n)^{1-\alpha}(c+\mathrm{o}(1)) = l\ln 2,$$

$$(\ln n)^{\alpha/(1-\alpha)}\ln\ln n = \left(\frac{l\ln 2}{c+\mathrm{o}(1)}\right)^{1/(1-\alpha)},$$

$$e^{\frac{\alpha}{1-\alpha}\ln\ln n}\left(\frac{\alpha}{1-\alpha}\ln\ln n\right) = \frac{\alpha}{1-\alpha}\left(\frac{l\ln 2}{c+\mathrm{o}(1)}\right)^{1/(1-\alpha)},$$

$$\frac{\alpha}{1-\alpha}\ln\ln n = W\left(\frac{\alpha}{1-\alpha}\left(\frac{l\ln 2}{c+\mathrm{o}(1)}\right)^{1/(1-\alpha)}\right),$$

где W — функция Ламберта. Далее, имеем

$$\ln n = \exp\left(\frac{1-\alpha}{\alpha}W\left(\frac{\alpha}{1-\alpha}\left(\frac{l\ln 2}{c+o(1)}\right)^{1/(1-\alpha)}\right)\right).$$

По определению W-функции $e^{W(x)} = x/W(x)$. Кроме того,

$$W(x) = \ln x - \ln\ln x + \mathrm{o}(1) = \ln x\left(1 - \frac{\ln\ln x}{\ln x} + \mathrm{o}(1)\right) = \ln(x)(1-\mathrm{o}(1)).$$

Соответственно получаем

$$\ln n = \left(\frac{\dfrac{\alpha}{1-\alpha}\left(\dfrac{l\ln 2}{c+\mathrm{o}(1)}\right)^{1/(1-\alpha)}}{W\left(\dfrac{\zeta}{1-\alpha}\left(\dfrac{l\ln 2}{c+\mathrm{o}(1)}\right)^{1/(1-\alpha)}\right)}\right)^{(1-\alpha)/\alpha} =$$

$$= \left(\frac{\frac{\alpha}{1-\alpha}\left(\frac{l\ln 2}{c+\mathrm{o}(1)}\right)^{1/(1-\alpha)}}{\ln\left(\frac{\alpha}{1-\alpha}\left(\frac{l\ln 2}{c+\mathrm{o}(1)}\right)^{1/(1-\alpha)}\right)(1-\mathrm{o}(1))} \right)^{(1-\alpha)/\alpha} = l^{1/\alpha}\left(\frac{\ln 2}{c+\mathrm{o}(1)}\right)^{1/\alpha}\left(\frac{\alpha}{1-\alpha}\right)^{(1-\alpha)/\alpha} \times$$

$$\times \left(\left(\ln\frac{\alpha}{1-\alpha} + \frac{1}{1-\alpha}\ln l + \frac{1}{1-\alpha}\ln\left(\frac{\ln 2}{c+\mathrm{o}(1)}\right)\right)(1-\mathrm{o}(1))\right)^{-(1-\alpha)/\alpha} =$$

$$= l^{1/\alpha}(\ln l)^{-(1-\alpha)/\alpha}\left(\frac{\ln 2}{c+\mathrm{o}(1)}\right)^{1/\alpha}\left(\frac{\alpha}{1-\mathrm{o}(1)}\right)^{(1-\alpha)/\alpha}.$$

3) $n^{c+\mathrm{o}(1)} = 2^l$, $\ln n = l\ln 2/(c+\mathrm{o}(1))$.

Для определения размера группы в виде $\lg n$ необходимо поделить все полученные величины на $\ln 2$. ■

Значение $s_l(\alpha, c)$ при $0 < \alpha < 1$ получено аппроксимацией W-функции Ламберта с использованием только первых двух членов разложения функции в ряд и отбрасыванием остальных. Эту функцию можно вычислить с произвольной точностью [36]. Методы для вычисления W-функции есть в составе многих систем компьютерной алгебры, например Maxima или PARI/GP. Поэтому значение $s_l(\alpha, c)$ при $0 < \alpha < 1$ более точно можно вычислить по формуле

$$s_l(\alpha, c) = \exp\left(\frac{1-\alpha}{\alpha}W\left(\frac{\alpha}{1-\alpha}\left(\frac{l\ln 2}{c+\mathrm{o}(1)}\right)^{1/(1-\alpha)}\right) - \ln\ln 2\right). \tag{9}$$

Следствие 1. Пусть C — гиперэллиптическая кривая рода g над полем $\mathbb{F}_q$; l — уровень безопасности; G — подгруппа порядка r якобиана кривой $Jac_C(\mathbb{F}_q)$; $\rho = g\ln q/\ln r$ — константа. Предположим, что сложность решения задачи ВДЛ в группе G выражается в виде $L_r(\alpha, c)$ битовых операций. Тогда размер поля $\lg q$, необходимый для обеспечения уровня безопасности l, равен $\dfrac{\rho}{g}s_l(\alpha, c)$.

В случае квазиполиномиальной сложности решения задачи ВДЛ имеет место

Утверждение 4. Пусть сложность решения задачи ВДЛ в группе G порядка n квазиполиномиальна — $2^{\mathrm{O}((\ln\ln n)^c)}$. Тогда размер группы $\lg n$, необходимый для обеспечения уровня безопасности l, равен $e^{\mathrm{O}(l^{1/c})}$.

Табл. 3 содержит необходимые размеры групп для обеспечения уровня безопасности l. Значения, обозначенные $l_1(G)$, рассчитаны по теореме 5 и утверждению 4; $l_2(G)$ обозначает размер группы G, найденный, где применимо, по более точной формуле (9). При этом значение ρ считалось равным 1; сложность алгоритмов взята из табл. 1 и 2.

Размер базового поля, требуемый для обеспечения уровня безопасности l, вычисленный по следствию 1, приведён в табл. 4; используемые сокращения: ЭК — эллиптическая кривая; ГЭК — гиперэллиптическая кривая.

При вычислении функции спаривания выполняются операции в поле $\mathbb{F}_{q^k}$; в общем случае чем больше это поле, тем ниже эффективность вычисления функции спаривания. Размер поля также можно вычислить по теореме 5 либо использовать следующее эквивалентное утверждение.

Утверждение 5. Пусть C — ГЭК рода g над полем $\mathbb{F}_q$, k — её степень вложения, $L_{q^k}(\alpha, c_1)$ и $L_r(\beta, c_2)$ — сложности решения задачи ВДЛ в конечном поле $\mathbb{F}_{q^k}$ и якобиане кривой $Jac_C(\mathbb{F}_q)$ соответственно, причём $0 < \alpha \leqslant \beta$. Если уровни безопасности

Таблица 3

Размеры групп для обеспечения уровня безопасности l

Группа G	Алгоритм	Размер группы, $l_1(G)$	Размер группы, $l_2(G)$
Любая	Pollard	$2\,l$	—
$E(\mathbb{F}_q)$, специальная	Semaev	$0{,}252\,l^2/\ln l$	$1{,}44\exp(W(0{,}350\,l^2))$
$Jac_C(\mathbb{F}_q)$, $g > O(\ln q)$	EG	$0{,}173\,l^2/\ln l$	$1{,}44\exp(W(0{,}240\,l^2))$
$Jac_C(\mathbb{F}_q)$, $g > \mathrm{O}(\ln^2 q)$	EGT	$0{,}007\,l^3/\ln^2 l$	$1{,}44\exp(2W(0{,}108\,l^{3/2}))$
$Jac_C(\mathbb{F}_q)$, $g = 3$	GTTD	$2{,}250\,l$	—
$Jac_C(\mathbb{F}_q)$, $g = 4$	GTTD	$2{,}666\,l$	—
$Jac_C(\mathbb{F}_q)$, $g = 5$	GTTD	$3{,}125\,l$	—
$\mathbb{F}_q^\times$, малое p	Joux [9]	$0{,}005\,l^4/(c^4\ln^3 l)$	$1{,}44\exp(3W(0{,}204\,l^{4/3})/c^{4/3})$
$\mathbb{F}_q^\times$, малое p	BGJT [10]	$\exp(\mathrm{O}(\sqrt{l}))$	—
$\mathbb{F}_q^\times$, среднее p	exTNFS	$0{,}010\,l^3/\ln^2 l$	$1{,}44\exp(2W(0{,}124\,l^{3/2}))$
$\mathbb{F}_q^\times$, большое p	MNFS [23]	$0{,}007\,l^3/\ln^2 l$	$1{,}44\exp(2W(0{,}110\,l^{3/2}))$

Таблица 4

Размеры базового поля ГЭК для уровня безопасности l

Кривая C	Алгоритм	Размер поля, $\lg q$
ЭК	Pollard	$2l$
ЭК, специальная	Semaev	$0{,}252\,l^2/\ln l$
ГЭК рода 2	Pollard	l
ГЭК рода 3	GTTD	$0{,}75\,l$
ГЭК рода 4	GTTD	$0{,}666\,l$
ГЭК рода 5	GTTD	$0{,}625\,l$

в якобиане кривой и поле $\mathbb{F}_{q^k}$ сбалансированы, то размер расширенного поля $\ln q^k$ равен

$$\frac{(\ln r)^{\beta/\alpha}}{(\ln\ln r)^{\beta/\alpha-1}}c,$$

где c — константа из теоремы 4.

В [13] для балансирования уровней безопасности в случае эллиптических кривых используется следующий метод. Фиксируется безопасный размер подгруппы точек эллиптической кривой b_1 и безопасный размер конечного поля b_2, затем для балансирования уровней безопасности используется соотношение $b_2/b_1 = \rho k$. В случае гиперэллиптических кривых необходимо учитывать род кривой, и соотношение принимает следующий вид: $b_2/b_1 = (\rho/g)k$.

Для получения значений b_1, b_2 могут использоваться рекомендации из [35, 37] или их можно приближенно вычислить, используя теорему 5. В последнем случае получаем

$$k = \frac{g}{\rho}\frac{s_l(\alpha, c_1)}{s_l(\beta, c_2)}.$$

Для расчёта границ сбалансированной степени вложения для заданного уровня безопасности l используем формулу из теоремы 4 с подстановкой вместо $\ln r$ требуемого размера группы для данной кривой из табл. 3. Для удобства значения k и q^k представлены в виде $al^{\lambda_1}\ln^{\lambda_2} l = \mathrm{O}(l^{\lambda_1}\ln^{\lambda_2} l) = \tilde{\mathrm{O}}(l^{\lambda_1})$, где a, λ_1, λ_2 — константы. Члены меньших порядков опускаются. Расчёт этих констант для различных полей и сбалансированного значения k представлен в табл. 5.

Заметим, что в случае малой характеристики существует теоретический квазиполиномиальный алгоритм. Из утверждения 4 следует, что в этом случае размер группы, необходимый для обеспечения безопасности, растёт экспоненциально от l. В то же время, по утверждению авторов алгоритма [10], константа в оценке сложности алгоритма большая и поэтому на практике он неприменим. Для табл. 5 были выбраны практические алгоритмы, но необходимо учитывать, что безопасность в полях малой характеристики остаётся под вопросом.

Таблица 5

Сбалансированная степень вложения для различных ГЭК и полей

g	ρ	p	Алгоритмы	Степень вложения, k			Размер поля $\mathbb{F}_{q^k}$		
				a	λ_1	λ_2	a	λ_1	λ_2
1	1	2	Joux [9] / Semaev [26]	0,0025 – 0,1648	2	−2	0,0006 – 0,0416	4	−3
		Малое	Joux [9] / Pollard	0,0026 – 0,1665	3	−3	0,0052 – 0,3330	4	−3
		Среднее	ExTNFS / Pollard	0,0050 – 0,0450	2	−2	0,0100 – 0,0901	3	−2
		Большое	MNFS / Pollard	0,0038 – 0,0349	2	−2	0,0077 – 0,0699	3	−2
	2	Малое	Joux [9] / Pollard	0,0013 – 0,0832	3	−3	0,0052 – 0,3330	4	−3
		Среднее	ExTNFS / Pollard	0,0025 – 0,0225	2	−2	0,0100 – 0,0901	3	−2
		Большое	MNFS / Pollard	0,0019 – 0,0174	2	−2	0,0077 – 0,0699	3	−2
2	1	Малое	Joux [9] / Pollard	0,0052 – 0,3330	3	−3	0,0052 – 0,3330	4	−3
		Среднее	ExTNFS / Pollard	0,0100 – 0,0901	2	−2	0,0100 – 0,0901	3	−2
		Большое	MNFS / Pollard	0,0077 – 0,0699	2	−2	0,0077 – 0,0699	3	−2
	2	Малое	Joux [9] / Pollard	0,0026 – 0,1665	3	−3	0,0052 – 0,3330	4	−3
		Среднее	ExTNFS / Pollard	0,0050 – 0,0450	2	−2	0,0100 – 0,0901	3	−2
		Большое	MNFS / Pollard	0,0038 – 0,0349	2	−2	0,0077 – 0,0699	3	−2
3	1	Малое	Joux [9] / GTTD [27]	0,0069 – 0,4440	3	−3	0,0052 – 0,3330	4	−3
		Среднее	ExTNFS / GTTD [27]	0,0133 – 0,1201	2	−2	0,0100 – 0,0901	3	−2
		Большое	MNFS / GTTD [27]	0,0103 – 0,0932	2	−2	0,0077 – 0,0699	3	−2
	2	Малое	Joux [9] / Pollard	0,0039 – 0,2497	3	−3	0,0052 – 0,3330	4	−3
		Среднее	ExTNFS / Pollard	0,0075 – 0,0675	2	−2	0,0100 – 0,0901	3	−2
		Большое	MNFS / Pollard	0,0058 – 0,0524	2	−2	0,0077 – 0,0699	3	−2

5. Сравнение кривых

Пусть l — некоторый фиксированный уровень безопасности. При сравнении учитываются следующие параметры — размер группы, размер базового поля, размер расширенного поля и сбалансированная степень вложения.

5.1. Размер группы

От размера группы зависит скорость вычисления спаривания, так как алгоритм Миллера, используемый для вычисления спаривания, в общем случае имеет сложность $\mathrm{O}(\log r)$.

Размер группы, необходимый для обеспечения уровня безопасности l, зависит от используемого алгоритма решения задачи ВДЛ в данной группе. Расчёт требуемых размеров групп представлен в табл. 3. Эллиптические кривые и гиперэллиптические кривые рода 2 имеют преимущество перед кривыми больших родов, так как требуют меньший размер группы.

5.2. Размер базового поля

Кривые больших родов позволяют использовать меньший размер базового поля (см. табл. 4). Размер поля влияет на эффективность вычислений в якобиане кривой и на размер параметров и ключей криптосистемы. Однако, как замечено в [38], существуют методы сжатия точек эллиптической кривой, позволяющие достичь размеров

параметров, аналогичных гиперэллиптическим кривым рода 2. Для кривых рода $g \geqslant 2$ подобные методы не разработаны.

5.3. Сложность группового закона

Для эллиптических кривых и гиперэллиптических кривых рода 2 существуют быстрые явные формулы для вычисления группового закона. В общем случае чем больше род кривой, тем сложнее групповой закон на кривой. Для кривых рода $g > 2$ по утверждению 2 сложность вычисления группового закона растёт как $\mathrm{O}(g^2)$ операций в базовом поле или $\mathrm{O}(g^2 \log^2 q)$ битовых операций.

С другой стороны, кривые большего рода требуют меньший размер базового поля для обеспечения аналогичного уровня безопасности (см. табл. 4).

5.4. Размер расширенного поля

Сбалансированная степень вложения определяет размер расширенного поля, над которым вычисляется функция спаривания (см. табл. 5). В общем случае чем больше это поле, тем менее эффективно вычисляется функция спаривания. Однако существуют методы, позволяющие минимизировать необходимое число операций в поле при вычислении спаривания. Кроме того, эффективность вычисления функции спаривания может быть улучшена с помощью использования кручений (кривых, изоморфных данной над $\overline{F}_q$) при их наличии.

5.5. Алгоритмы генерации и ограничения на параметры

Так как для случайно выбранной эллиптической или гиперэллиптической кривой вероятность того, что её степень вложения мала, незначительна, кривые, подходящие для использования в криптографии на билинейных спариваниях, требуют специальных алгоритмов построения. Исключение составляют суперсингулярные кривые, которые всегда имеют малую степень вложения для любого рода (для эллиптических кривых $k \leqslant 6$ [1], для кривых больших родов границы рассчитаны в работе [39]).

Алгоритмы генерации наиболее разработаны для эллиптических кривых, в меньшей степени — для ГЭК рода 2, в отдельных случаях — для рода 3 и практически полностью не разработаны для рода 4 и больше. Классификация эллиптических кривых, подходящих для использования в криптографии на билинейных спариваниях, а также методы генерации таких кривых есть в [13]. Для сравнения из данной работы были выбраны кривые с наименьшими значениями ρ (см. [13, Table 8.2]).

В случае суперсингулярных эллиптических кривых возможны только степени вложения $k = 1, 2, 3, 4, 6$, причём значения $k = 4$ и 6 возможны только в случае $p = 2$ и 3 соответственно. Известно также, что при $k = 1$ минимальное ρ-значение равно 2.

В случае обычных эллиптических кривых существует несколько семейств кривых, подходящих для криптографии на спариваниях, — MNT-кривые, кривые Кокса — Пинча и др. Алгоритмы генерации таких кривых позволяют строить кривые с произвольной степенью вложения.

Хороший обзор методов генерации гиперэллиптических кривых для криптографии на билинейных спариваниях приведён в [14]. В отличие от эллиптических кривых, в настоящее время не разработаны методы генерации, позволяющие получать ГЭК со значением $\rho < 2$.

В случае ГЭК рода 2 первый алгоритм для генерации обычных ГЭК рода 2 с произвольной степенью вложения предложен Фриманом [40]. Алгоритм основан на обобщении метода Кокса — Пинча и позволяет генерировать кривые с $\rho \approx 8$. Позже было предложено обобщение метода Брезинга — Венга [41], позволяющее в частных случаях получать кривые с меньшим ρ-значением.

В случае кривых с разделённым якобианом существуют методы [42 – 45] для генерации кривых с $\rho \approx 4$. В частных случаях для определённых значений k возможны меньшие ρ-значения; например, в [45] для $k = 3, 4, 6, 12$ получены кривые с $\rho \approx 2$.

Суперсингулярные гиперэллиптические кривые рода 2 имеют степень вложения $k \leqslant 12$. При этом степень вложения $k = 12$ возможна только для кривой над полем характеристики 2, а степень вложения $k = 4$ — только для характеристики 3. Суперсингулярные кривые в настоящее время являются единственным известным классом ГЭК рода 2, для которых достигается значение $\rho \approx 1$.

В случае кривых рода 3 в работах [46, 41] предложены методы, позволяющие получить кривые в отдельных случаях с $\rho \approx 12$ при $k = 7$, $\rho \approx 15$ при $k = 9, 18$, $\rho \approx 8$ при $k = 13$.

5.6. Диапазон уровней безопасности и результаты

Если степень вложения k кривой C фиксирована, то, начиная с некоторого уровня безопасности l, эта степень становится меньше сбалансированного значения k, а такая кривая — непригодной для криптографии с уровнем безопасности больше l.

Для расчёта такого значения l введём следующий метод. Сначала сбалансированное значение k выражается в виде $al^{\lambda_1}(\ln l)^{\lambda_2}$. Значения $(a, \lambda_1, \lambda_2)$ можно брать из табл. 5 либо вычислять, используя теоремы 4 и 5. Далее, из равенства $k = al^{\lambda_1}(\ln l)^{\lambda_2}$ выражаем l и получаем формулу

$$l = \exp\left(\frac{\lambda_2}{\lambda_1} W\left(\frac{\lambda_1}{\lambda_2}\left(\frac{k}{a}\right)^{1/\lambda_2}\right)\right).$$

В настоящее время для криптографии требуется $r > 2^{160}$ и в общем случае все алгоритмы ВДЛ в якобиане кривых экспоненциальны, поэтому введём следующую более точную теорему для данного частного случая.

Теорема 6. Пусть C — ГЭК рода 2; $G \subseteq Jac_C(\mathbb{F}_q)$ — подгруппа простого порядка $r > e^e$; k — степень вложения кривой; $L_{q^k}(\alpha, c_1), L_r(\beta, c_2)$ — сложности решения задачи ВДЛ в конечном поле $\mathbb{F}_{q^k}$ и группе G соответственно. Тогда если $\beta = 1$, то кривая удовлетворяет уровню безопасности

$$l = \frac{c_2}{\ln 2}\frac{W_-(d)}{d},$$

где $d = -(gc/k\rho)^{\alpha/(\beta-\alpha)}$ и c — константа из теоремы 3.

Доказательство. По теореме 3 получаем сбалансированную степень вложения

$$k = \frac{g}{\rho} c \left(\frac{\ln r}{\ln\ln r}\right)^{\beta/\alpha-1}.$$

Выразим из данного выражения $\ln r$:

$$\frac{k\rho}{gc} = (\ln r)^{\beta/\alpha-1}(\ln\ln r)^{-(\beta/\alpha-1)} = e^{(\beta/\alpha-1)\ln\ln r}(\ln\ln r)^{-(\beta/\alpha-1)} = \left(e^{-\ln\ln r}\ln\ln r\right)^{-(\beta/\alpha-1)},$$

$$-\left(\frac{gc}{k\rho}\right)^{1/(\beta/\alpha-1)} = e^{-\ln\ln r}(-\ln\ln r), \quad \ln\ln r = -W(d), \quad \ln r = e^{-W(d)} = \frac{W(d)}{d}.$$

Здесь W-функция Ламберта определяется как решение уравнения $We^W = x$. На интервале $(-1/e, 0)$ уравнение имеет два вещественных решения, одно возрастает от -1 до 0 при $d \to 0$ и обозначается W_+; второе стремится к $-\infty$ при $d \to 0$.

Для W_+ имеем $1 < (W_+(d))/d < e$, то есть получаем $1 < \ln r < e$, что противоречит условию $r > e^e$. Поэтому остаётся решение W_-, и $\ln r = (W_-(d))/d$. Применяя теорему 5 для $\beta = 1$, получаем

$$l = \frac{c_2}{\ln 2} \ln r = \frac{c_2}{\ln 2} \frac{W_-(d)}{d}.$$

Теорема доказана. ■

Для сравнения были выбраны кривые со степенями вложения $1 \leqslant k \leqslant 10$ по следующим правилам:

1) для кривой или семейства кривых должны быть известны алгоритмы генерации и явные примеры построения, которые указываются в графе «Источник»;
2) выбираются кривые с минимальным ρ-значением среди всех кривых с одинаковой степенью вложения, родом и полем;
3) в случае, если ρ-значения совпадают, выбирается любая из кривых;
4) если требуются дополнительные ограничения или кривая имеет специальный вид, то дополнительно указываются кривые с минимальным ρ-значением без ограничений, а кривая помечается знаком «+»;
5) указываются кривые, уровень безопасности которых не удовлетворяет современным требованиям [47, 48] ($l < 80$). В этом случае кривая помечается знаком «*».

Результаты представлены в табл. 6; диапазон уровней безопасности рассчитан по теореме 6.

Заключение

В работе введена формула (теорема 4) для оценки границ сбалансированного значения степени вложения гиперэллиптических кривых.

Для гиперэллиптических кривых рода 1–3 вычислены текущие границы сбалансированной степени вложения для различных полей (табл. 5). Кривые со степенями вложения, выходящими за эти границы, либо не безопасны, либо их степень вложения превосходит необходимую для заданного уровня безопасности.

Для кривых с известными алгоритмами генерации выбраны семейства с наименьшими ρ-значениями и произведены вычисления диапазона уровней безопасности, которому они могут соответствовать (табл. 6), что позволило определить некоторые кривые, небезопасные для использования в настоящее время.

Таблица 6

Уровни безопасности для гиперэллиптических кривых

k	Кривая C	g	ρ	p	Источники	l
1	ЭК, полная	1	2	Среднее	[13, § 6.6]	23–98
1	ЭК, полная	1	2	Большое	[13, § 6.6]	27–115
2	ЭК, суперсингулярная	1	1	Среднее	[13, § 3.2]	23–98
2	ЭК, суперсингулярная	1	1	Большое	[13, § 3.2]	27–115
2	ГЭК, обычная$^+$	2	3	Большое	[45, Table 1]	36–147
2	ГЭК, обычная	2	8,135	Большое	[40, Ex. 1]	69–271
3	ЭК, суперсингулярная	1	1	Среднее	[13, § 3.3]	30–126
3	ЭК, суперсингулярная	1	1	Большое	[13, § 3.3]	36–147
3	ГЭК, обычная$^+$	2	2	Большое	[45, Ex. 24]	36–147
4	ЭК, полная	1	1,5	Среднее	[13, § 6.9]	48–193
4	ЭК, полная	1	1,5	Большое	[13, § 6.9]	57–225
4	ЭК, суперсингулярная*	1	1	2	[13, § 3.4]	6–48
4	ЭК, суперсингулярная*	1	2	2	[13, § 3.4]	9–65
4	ГЭК, суперсингулярная*	2	1	3	[14, § 3.4]	3–35
4	ГЭК, обычная$^+$	2	2	Большое	[45, Ex. 24]	43–176
4	ГЭК, обычная	2	8,139	Большое	[40, Ex. 4]	108–410
5	ЭК, полная	1	1,5	Среднее	[13, § 6.6]	56–221
5	ЭК, полная	1	1,5	Большое	[13, § 6.6]	66–258
5	ГЭК, обычная$^+$	2	3	Большое	[45, Table 1]	66–258
5	ГЭК, обычная	2	4	Большое	[41, 5.2]	79–307
6	ЭК, суперсингулярная*	1	1	3	[13, § 3.5]	7–57
6	ЭК, суперсингулярная*	1	2	3	[13, § 3.5]	11–77
6	ЭК, Скотта — Баретто	1	1,25	Среднее	[13, § 6.16]	56–221
6	ЭК, Скотта — Баретто	1	1,25	Большое	[13, § 6.16]	66–258
6	ГЭК, суперсингулярная*	2	1	Большое	[14, § 3.4]	5–42
6	ГЭК, обычная$^+$	2	2	Большое	[45, Ex. 24]	57–225
6	ГЭК, обычная	2	7,5	Большое	[41, Table 1]	132–497
7	ЭК, циклотомическая	1	1,33	Среднее	[13, § 6.20]	64–252
7	ЭК, циклотомическая	1	1,33	Большое	[13, § 6.20]	75–294
7	ГЭК, обычная$^+$	2	2,5	Большое	[45, Table 1]	73–283
7	ГЭК, обычная	3	12	Большое	[41, 5.7]	151–565
8	ЭК, полная	1	1,25	Среднее	[13, § 6.6]	67–263
8	ЭК, полная	1	1,25	Большое	[13, § 6.6]	79–307
8	ГЭК, обычная$^+$	2	3	Большое	[45, Ex. 25]	89–342
8	ГЭК, обычная	2	7,5	Большое	[41, Table 1]	157–589
9	ЭК, полная	1	1,33	Среднее	[13, § 6.6]	75–293
9	ЭК, полная	1	1,33	Большое	[13, § 6.6]	89–341
9	ГЭК, обычная$^+$	2	2,33	Большое	[45, Table 1]	81-315
9	ГЭК, обычная	3	15	Большое	[41, 5.7]	202–747
10	ЭК, полная	1	1,5	Среднее	[13, § 6.5]	87–336
10	ЭК, полная	1	1,5	Большое	[13, § 6.5]	102–391
10	ГЭК, обычная$^+$	2	3	Большое	[43, Th. 4]	102–391
10	ГЭК, обычная	2	6	Большое	[41, 5.4]	157–589

ЛИТЕРАТУРА

1. *Menezes A., Okamoto T., and Vanstone S. A.* Reducing elliptic curve logarithms to logarithms in a finite field // IEEE Trans. Inform. Theory. 1993. V. 39. No. 5. P. 1639–1646.
2. *Sakai R., Ohgishi K., and Kasahara M.* Cryptosystems Based on Pairing. Okinawa, Japan, 2000.
3. *Joux A.* A one round protocol for tripartite Diffie — Hellman // ANTS-IV. LNCS. 2000. V. 1838. P. 385–393.

4. *Boneh D. and Franklin M. K.* Identity-based encryption from the Weil pairing // CRYPTO 2001. LNCS. 2001. V. 2139. P. 213–229.
5. *Paterson K. G.* Cryptography from pairings // Advances in Elliptic Curve Cryptography. Cambridge University Press, 2005. P. 215–252.
6. *Hitt L.* On the minimal embedding field // LNCS. 2007. V. 4575. P. 294–301.
7. *Benger N., Charlemagne M., and Freeman D. M.* On the security of pairing-friendly Abelian varieties over non-prime fields // LNCS. 2009. V. 5671. P. 52–65.
8. *Galbraith S. D., Hess F., and Vercauteren F.* Aspects of pairing inversion // IEEE Trans. Inform. Theory. 2008. V. 54. No. 12. P. 5719–5728.
9. *Joux A.* A new index calculus algorithm with complexity $L(1/4 + o(1))$ in small characteristic // LNCS. 2014. V. 8282. P. 355–379.
10. *Barbulescu R., Gaudry P., Joux A., and Thomé E.* A heuristic quasi-polynomial algorithm for discrete logarithm in finite fields of small characteristic // LNCS. 2014. V. 8441. P. 1–16.
11. *Frey G. and Lange T.* Transfer of discrete logarithms // Handbook of Elliptic and Hyperelliptic Curve Cryptography. Chapman and Hall/CRC, 2005. P. 529–545.
12. *Miller V. S.* The Weil pairing, and its efficient calculation // J. Cryptology. 2004. V. 17. No. 4. P. 235–261.
13. *Freeman D., Scott M., and Teske E.* A taxonomy of pairing-friendly elliptic curves // J. Cryptology. 2010. V. 23. No. 2. P. 224–280.
14. *Balakrishnan J., Belding J., Chisholm S., et al.* Pairings on hyperelliptic curves. `http://arxiv.org/abs/0908.373`. 2009.
15. *Pollard J. M.* Monte Carlo methods for index computation (mod p) // Math. Comput. 1978. V. 78. No. 147. P. 918–924.
16. *Pohlig S. C. and Hellman M. E.* An improved algorithm for computing logarithms over GF(p) and its cryptographic significance (Corresp.) // IEEE Trans. Inform. Theory. 1978. V. 24. No. 1. P. 106–110.
17. *Joux A., Odlyzko A., and Pierrot C.* The past, evolving present, and future of the discrete logarithm // Open Problems in Mathematics and Computational Science. Springer International Publishing, 2014. P. 5–36.
18. *Adj G., Menezes A., Oliveira T., and Rodríguez-Henríquez F.* Computing discrete logarithms in $\mathbb{F}_{3^{6\cdot 137}}$ and $\mathbb{F}_{3^{6\cdot 163}}$ using Magma // LNCS. 2015. V. 9061. P. 3–22.
19. *Adj G., Menezes A., Oliveira T., and Rodríguez-Henríquez F.* Weakness of $\mathbb{F}_{3^{6\cdot 1429}}$ and $\mathbb{F}_{2^{4\cdot 3041}}$ for discrete logarithm cryptography // Finite Fields and Their Applications. 2015. V. 32. P. 148–170.
20. *Joux A. and Pierrot C.* Technical history of discrete logarithms in small characteristic finite fields — The road from subexponential to quasi-polynomial complexity // Des. Codes Cryptography. 2016. V. 78. No. 1. P. 73–85
21. *Pierrot C.* The multiple Number Field Sieve with conjugation and generalized Joux — Lercier methods // LNCS. 2015. V. 2056. P. 156–170.
22. *Kim T. and Barbulescu R.* Extended Tower Number Field Sieve: A New Complexity for Medium Prime Case. Cryptology ePrint Archive, Report 2015/1027. `http://ia.cr/2015/1027`.
23. *Pierrot C. and Barbulescu R.* The Multiple Number Field Sieve for Medium and High Characteristic Finite Fields. Cryptology ePrint Archive, Report 2014/147. `http://ia.cr/2014/147`.
24. *Frey G. and Rück H. G.* A remark concerning m-divisibility and the discrete logarithm in the divisor class group of curves // Math. Comput. 1994. V. 62. No. 206. P. 865–874.

25. *Gaudry P., Hess F., and Smart N. P.* Constructive and destructive facets of Weil descent on elliptic curves // J. Cryptology. 2002. V. 15. No. 1. P. 19–46.
26. *Semaev I. A.* New algorithm for the discrete logarithm problem on elliptic curves. Cryptology ePrint Archive, Report 2015/310. `http://ia.cr/2015/310`.
27. *Gaudry P., Thomé E., Thériault N., and Diem C.* A double large prime variation for small genus hyperelliptic index calculus // Math. Comput. 2007. V. 76. No. 257. P. 475–492.
28. *Enge A. and Gaudry P.* A general framework for subexponential discrete logarithm algorithms // Acta Arith. 2000. V. 102. P. 83–103.
29. *Enge A., Gaudry P., and Thomé E.* An $L(1/3)$ discrete logarithm algorithm for low degree curves // J. Cryptology. 2011. V. 24. No. 1. P. 24–41.
30. *Menezes A.* The elliptic curve logarithm problem // Elliptic Curve Public Key Cryptosystems. Boston: Springer US, 1993. P. 61–81.
31. *Lenstra A. K. L* Notation // Encyclopedia of Cryptography and Security. Springer, 2011. P. 709–710.
32. *Cantor D. G.* Computing in the Jacobian of a hyperelliptic curve // Math. Comput. 1987. V. 48. No. 177. P. 95–101.
33. *Jacobson M., Menezes A., and Stein A.* Hyperelliptic curves and cryptography // Fields Institute Communications. 2004. V. 41. P. 255–282.
34. *Galbraith S.* Pairings // Advances in Elliptic Curve Cryptography. Cambridge University Press, 2005. P. 183–212.
35. `http://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-57pt1r4.pdf` — NIST. Recommendation for Key Management, Part 1: General. 2016.
36. *Corless R. M., Gonnet G. H., Hare D. E. G., et al.* On the Lambert W function // Adv. Comput. Math. 1996. V. 5. No. 1. P. 329–359.
37. `http://www.ecrypt.eu.org/ecrypt2/documents/D.SPA.20.pdf` — ECRYPT II. Yearly Report on Algorithms and Keysizes. 2012.
38. *Galbraith S. D., Hess F., and Vercauteren F.* Hyperelliptic pairings // LNCS. 2007. V. 4575. P. 108–131.
39. *Galbraith S. D.* Supersingular curves in cryptography // LNCS. 2001. V. 2248. P. 495–513.
40. *Freeman D.* Constructing pairing-friendly genus 2 curves with ordinary Jacobians // LNCS. 2007. V. 4575. P. 152–176.
41. *Freeman D.* A Generalized Brezing — Weng Algorithm for Constructing Pairing-Friendly Ordinary Abelian Varieties. Cryptology ePrint Archive, Report 2008/155. `http://ia.cr/2008/155`.
42. *Kawazoe M. and Takahashi T.* Pairing-friendly hyperelliptic curves with ordinary Jacobians of type $y^2 = x^5 + ax$ // LNCS. 2008. V. 5209. P. 164–177.
43. *Kachisa E. J.* Generating more Kawazoe — Takahashi genus 2 pairing-friendly hyperelliptic curves // LNCS. 2010. V. 6487. P. 312–326.
44. *Freeman D. M. and Satoh T.* Constructing pairing-friendly hyperelliptic curves using Weil restriction // J. Number Theory. 2011. V. 131. No. 5. P. 959–983.
45. *Drylo R.* Constructing pairing-friendly genus 2 curves with split Jacobian // LNCS. 2012. V. 7668. P. 431–453.
46. *Freeman D., Stevenhagen P., and Streng M.* Abelian varieties with prescribed embedding degree // LNCS. 2008. V. 5011. P. 60–73.
47. *Lenstra A. K. and Verheul E. R.* Selecting cryptographic key sizes // J. Cryptology. 2001. V. 14. P. 255–293.
48. `www.keylength.com` — BlueKrypt: Cryptgraphic Key Length Recommendation. 2016.

REFERENCES

1. *Menezes A., Okamoto T., and Vanstone S. A.* Reducing elliptic curve logarithms to logarithms in a finite field. IEEE Trans. Inform. Theory. 1993, vol. 39, no. 5, pp. 1639–1646.
2. *Sakai R., Ohgishi K., and Kasahara M.* Cryptosystems Based on Pairing. Okinawa, Japan, 2000.
3. *Joux A.* A one round protocol for tripartite Diffie — Hellman. ANTS-IV, LNCS, 2000, vol. 1838, pp. 385–393.
4. *Boneh D. and Franklin M. K.* Identity-based encryption from the Weil pairing. CRYPTO 2001, LNCS, 2001, vol. 2139, pp. 213–229.
5. *Paterson K. G.* Cryptography from pairings. Advances in Elliptic Curve Cryptography. Cambridge University Press, 2005, pp. 215–252.
6. *Hitt L.* On the minimal embedding field. LNCS, 2007, vol. 4575, pp. 294–301.
7. *Benger N., Charlemagne M., and Freeman D. M.* On the security of pairing-friendly Abelian varieties over non-prime fields. LNCS, 2009, vol. 5671, pp. 52–65.
8. *Galbraith S. D., Hess F., and Vercauteren F.* Aspects of pairing inversion. IEEE Trans. Inform. Theory. 2008, vol. 54, no. 12, pp. 5719–5728.
9. *Joux A.* A new index calculus algorithm with complexity $L(1/4+\mathrm{o}(1))$ in small characteristic. LNCS, 2014, vol. 8282, pp. 355–379.
10. *Barbulescu R., Gaudry P., Joux A., and Thomé E.* A heuristic quasi-polynomial algorithm for discrete logarithm in finite fields of small characteristic. LNCS, 2014, vol. 8441, pp. 1–16.
11. *Frey G. and Lange T.* Transfer of discrete logarithms. Handbook of Elliptic and Hyperelliptic Curve Cryptography. Chapman and Hall/CRC, 2005, pp. 529–545.
12. *Miller vol. S.* The Weil pairing, and its efficient calculation. J. Cryptology, 2004, vol. 17, no. 4, pp. 235–261.
13. *Freeman D., Scott M., and Teske E.* A taxonomy of pairing-friendly elliptic curves. J. Cryptology, 2010, vol. 23, no. 2, pp. 224–280.
14. *Balakrishnan J., Belding J., Chisholm S., et al.* Pairings on hyperelliptic curves. `http://arxiv.org/abs/0908.373`, 2009.
15. *Pollard J. M.* Monte Carlo methods for index computation $\pmod p$. Math. Comput., 1978, vol. 78, no. 147, pp. 918–924.
16. *Pohlig S. C. and Hellman M. E.* An improved algorithm for computing logarithms over $\mathrm{GF}(p)$ and its cryptographic significance (Corresp.). IEEE Trans. Inform. Theory, 1978, vol. 24, no. 1, pp. 106–110.
17. *Joux A., Odlyzko A., and Pierrot C.* The past, evolving present, and future of the discrete logarithm. Open Problems in Mathematics and Computational Science. Springer International Publishing, 2014, pp. 5–36.
18. *Adj G., Menezes A., Oliveira T., and Rodríguez-Henríquez F.* Computing discrete logarithms in $\mathbb{F}_{3^{6\cdot 137}}$ and $\mathbb{F}_{3^{6\cdot 163}}$ using Magma. LNCS, 2015, vol. 9061. pp. 3–22.
19. *Adj G., Menezes A., Oliveira T., and Rodríguez-Henríquez F.* Weakness of $\mathbb{F}_{3^{6\cdot 1429}}$ and $\mathbb{F}_{2^{4\cdot 3041}}$ for discrete logarithm cryptography. Finite Fields and Their Applications, 2015, vol. 32, pp. 148–170.
20. *Joux A. and Pierrot C.* Technical history of discrete logarithms in small characteristic finite fields — The road from subexponential to quasi-polynomial complexity. Des. Codes Cryptography, 2016, vol. 78, no. 1, pp. 73–85
21. *Pierrot C.* The multiple Number Field Sieve with conjugation and generalized Joux — Lercier methods. LNCS, 2015, vol. 2056, pp. 156–170.

22. *Kim T. and Barbulescu R.* Extended Tower Number Field Sieve: A New Complexity for Medium Prime Case. Cryptology ePrint Archive, Report 2015/1027. http://ia.cr/2015/1027.
23. *Pierrot C. and Barbulescu R.* The Multiple Number Field Sieve for Medium and High Characteristic Finite Fields. Cryptology ePrint Archive, Report 2014/147. http://ia.cr/2014/147.
24. *Frey G. and Rück H. G.* A remark concerning m-divisibility and the discrete logarithm in the divisor class group of curves. Math. Comput., 1994, vol. 62, no. 206, pp. 865–874.
25. *Gaudry P., Hess F., and Smart N. P.* Constructive and destructive facets of Weil descent on elliptic curves. J. Cryptology, 2002, vol. 15, no. 1, pp. 19–46.
26. *Semaev I. A.* New algorithm for the discrete logarithm problem on elliptic curves. Cryptology ePrint Archive, Report 2015/310. http://ia.cr/2015/310.
27. *Gaudry P., Thomé E., Thériault N., and Diem C.* A double large prime variation for small genus hyperelliptic index calculus. Math. Comput., 2007, vol. 76, no. 257, pp. 475–492.
28. *Enge A. and Gaudry P.* A general framework for subexponential discrete logarithm algorithms. Acta Arith., 2000, vol. 102, pp. 83–103.
29. *Enge A., Gaudry P., and Thomé E.* An $L(1/3)$ discrete logarithm algorithm for low degree curves. J. Cryptology, 2011, vol. 24, no. 1, pp. 24–41.
30. *Menezes A.* The elliptic curve logarithm problem. Elliptic Curve Public Key Cryptosystems. Boston: Springer US, 1993, pp. 61–81.
31. *Lenstra A. K.* L Notation. Encyclopedia of Cryptography and Security. Springer, 2011, pp. 709–710.
32. *Cantor D. G.* Computing in the Jacobian of a hyperelliptic curve. Math. Comput., 1987, vol. 48, no. 177, pp. 95–101.
33. *Jacobson M., Menezes A., and Stein A.* Hyperelliptic curves and cryptography. Fields Institute Communications, 2004, vol. 41, pp. 255–282.
34. *Galbraith S.* Pairings. Advances in Elliptic Curve Cryptography. Cambridge University Press, 2005, pp. 183–212.
35. http://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-57pt1r4.pdf — NIST. Recommendation for Key Management, Part 1: General. 2016.
36. *Corless R. M., Gonnet G. H., Hare D. E. G., et al.* On the Lambert W function. Adv. Comput. Math., 1996, vol. 5, no. 1, pp. 329–359.
37. http://www.ecrypt.eu.org/ecrypt2/documents/D.SPA.20.pdf — ECRYPT II. Yearly Report on Algorithms and Keysizes, 2012.
38. *Galbraith S. D., Hess F., and Vercauteren F.* Hyperelliptic pairings. LNCS, 2007, vol. 4575, pp. 108–131.
39. *Galbraith S. D.* Supersingular curves in cryptography. LNCS, 2001, vol. 2248, pp. 495–513.
40. *Freeman D.* Constructing pairing-friendly genus 2 curves with ordinary Jacobians. LNCS, 2007, vol. 4575, pp. 152–176.
41. *Freeman D.* A Generalized Brezing — Weng Algorithm for Constructing Pairing-Friendly Ordinary Abelian Varieties. Cryptology ePrint Archive, Report 2008/155. http://ia.cr/2008/155.
42. *Kawazoe M. and Takahashi T.* Pairing-friendly hyperelliptic curves with ordinary Jacobians of type $y^2 = x^5 + ax$. LNCS, 2008, vol. 5209, pp. 164–177.
43. *Kachisa E. J.* Generating more Kawazoe — Takahashi genus 2 pairing-friendly hyperelliptic curves. LNCS, 2010, vol. 6487, pp. 312–326.
44. *Freeman D. M. and Satoh T.* Constructing pairing-friendly hyperelliptic curves using Weil restriction. J. Number Theory, 2011, vol. 131, no. 5, pp. 959–983.

45. *Drylo R.* Constructing pairing-friendly genus 2 curves with split Jacobian. LNCS, 2012, vol. 7668, pp. 431–453.
46. *Freeman D., Stevenhagen P., and Streng M.* Abelian varieties with prescribed embedding degree. LNCS, 2008, vol. 5011, pp. 60–73.
47. *Lenstra A. K. and Verheul E. R.* Selecting cryptographic key sizes. J. Cryptology, 2001, vol. 14, pp. 255–293.
48. `www.keylength.com` — BlueKrypt: Cryptgraphic Key Length Recommendation, 2016.

МАТЕМАТИЧЕСКИЕ МЕТОДЫ СТЕГАНОГРАФИИ

УДК 519.7

ПОВЫШЕНИЕ ЭФФЕКТИВНОСТИ МЕТОДОВ СТЕГОАНАЛИЗА ПРИ ПОМОЩИ ПРЕДВАРИТЕЛЬНОЙ ФИЛЬТРАЦИИ КОНТЕЙНЕРОВ[1]

В. А. Монарёв*, А. И. Пестунов**

*Институт вычислительных технологий СО РАН, г. Новосибирск, Россия
**Новосибирский государственный университет экономики и управления, г. Новосибирск, Россия

Предлагается новый подход к стегоанализу, названный «предварительной фильтрацией», который заключается в том, чтобы перед финальным обнаружением добавить этап отбора «хороших» контейнеров, наличие/отсутствие внедрённой информации в которых может быть определено более достоверно, чем во всём множестве. При этом размер данного подмножества, точнее, его доля по отношению ко всему контрольному множеству может рассматриваться как дополнительная характеристика метода стегоанализа. Предлагаются три конкретных метода для предварительной фильтрации, которые названы «наивный метод», «простая классификация» и «комбинированная классификация». Приводятся результаты экспериментов по предварительной фильтрации изображений из известного множества BOSSbase v1.01.

Ключевые слова: *стегоанализ, ошибка обнаружения, адаптивное внедрение, HUGO, ансамблевый классификатор.*

DOI 10.17223/20710410/32/6

ENHANCING STEGANALYSIS ACCURACY VIA TENTATIVE FILTERING OF STEGO-CONTAINERS

V. A. Monarev*, A. I. Pestunov**

*Institute of Computational Technologies SB RAS, Novosibirsk, Russia
**Novosibirsk State University of Economics and Management, Novosibirsk, Russia

E-mail: viktor.monarev@gmail.com, pestunov@gmail.com

We introduce a new approach to steganalysis called “the tentative filtering” and consisting in inserting an additional filtering phase before the final classification in order to select those containers where stego-information can be reliably detected. The size of this “good” subset of containers can be considered as an additional characteristic of the detector. We introduce three methods for implementing the tentative filtering: the naive method, the simple classification, and the combined classification. The experiments demonstrate that it is possible to select about 35 % of BOSSbase v1.01 images, for which HUGO 0.4 bpp is detected with the error less than 0.003, while the error over the whole set is 0.141. It is also demonstrated that it is possible to select

[1]Работа поддержана грантом РФФИ, проект № 14-01-31484 (мол_а).

about 5 % images, for which HUGO 0.1 bpp is detected with the error less than 0.05, while the whole set gives the error 0.37 (which is not quite a reliable detection).

Keywords: *steganalysis, detection error, image features, SRM, adaptive steganography, HUGO, ensemble classifier.*

Введение

Вычисление ошибки обнаружения (detection error) по формуле $P_{\mathrm{E}} = (P_{\mathrm{FA}} + P_{\mathrm{MD}})/2$, где P_{FA} — вероятность ложного срабатывания (False Alarm), а P_{MD} — вероятность пропущенного обнаружения (Missed Detection), в последнее время является одним из наиболее распространённых подходов к оценке точности методов стегоанализа [1 – 5].

Проблема заключается в том, что на практике статистическая модель контейнеров редко бывает известной и ошибку обнаружения не удаётся вычислить аналитически. Обычно она вычисляется экспериментально и равняется отношению числа контейнеров, на которых метод стегоанализа отработал правильно, к числу всех контейнеров в контрольном множестве. Для этой цели специалисты по стегоанализу используют известные стандартизованные множества контейнеров, например BOSSbase [6, 7], BOWS2 [8] или NRCS [9]. Тем не менее ошибка обнаружения, вычисленная по различным множествам, может отличаться из-за их специфичных свойств (наличия шумов, степени сжатия и пр.) [10]. Более того, в рамках одного множества контейнеров могут выделяться подмножества, свойства которых различаются, что приводит к различиям в ошибках обнаружения, если вычислять их по этим подмножествам отдельно.

В настоящей работе предлагается новый подход к стегоанализу, который назван «предварительная фильтрация». Идея подхода заключается в том, чтобы перед финальным обнаружением добавить этап отбора тех контейнеров (назовём их «хорошими»), наличие/отсутствие внедрённой информации в которых может быть определено более достоверно, чем во всём множестве. Другими словами, ошибка обнаружения, вычисленная по отобранному подмножеству, будет ниже, чем ошибка, вычисленная по всему контрольному множеству. При этом размер данного подмножества, точнее, его доля по отношению ко всему контрольному множеству может рассматриваться как дополнительная характеристика метода стегоанализа. Аналогичная ситуация имеет место в криптографии при разработке атак в предположении использования слабых ключей, когда размер множества слабых ключей является одним из показателей эффективности атаки [11, 12]. Предварительная фильтрация позволит не только снизить ошибку обнаружения, но и более тонко оценивать точность методов стегоанализа, поскольку появляется возможность выбирать контейнеры, свойства которых подходят для заданного метода стегоанализа.

Предварительную фильтрацию следует рассматривать как общий подход к стегоанализу, поэтому конкретных методов в рамках этого подхода может быть разработано довольно много. В данной работе предлагается три возможных варианта: наивный метод, простая классификация и комбинированная классификация, являющаяся комбинацией первых двух методов. Эксперименты показали, что предварительная фильтрация позволяет выбрать порядка 35 % изображений из известного множества BOSSbase v1.01 [7], для которых метод адаптивной стеганографии HUGO 0,4 битов на пиксель (б/п) обнаруживается с ошибкой менее 0,003, в то время как ошибка, вычисленная по всему множеству, составляет 0,141. Показано также, что из всего множества можно выбрать порядка 5 % изображений, для которых HUGO 0,1 б/п определяется с ошибкой менее чем 0,05, тогда как ошибка по всему множеству составляет 0,37.

1. Предварительные замечания и обозначения

Предлагаемый подход (и методы в его рамках) может применяться к любым контейнерам, но поскольку все эксперименты проводились на изображениях, то во избежание разночтений далее вместо термина «контейнер» везде используется «изображение».

1.1. Задача бинарной классификации в стеганографии

Задача бинарной классификации заключается в том, чтобы отнести заданное изображение к одному из двух классов — пустое или заполненное, причём стегоаналитик действует по следующему сценарию [3]:

1) имеет доступ к изображениям, которые обладают статистическими свойствами, схожими с теми, которые используются для внедрения информации;
2) знает алгоритм внедрения и точный размер внедряемого сообщения (обычно он измеряется в битах на пиксель);
3) знает, какой объект он должен исследовать.

В рамках настоящей работы не затрагивается количественный стегоанализ [13, 14], когда стегоаналитик не знает размера внедряемого сообщения.

Современные подходы к решению задачи бинарной классификации состоят из двух основных этапов: выделения признаков (feature extraction) из изображения и непосредственно классификации [3]. При этом предполагается, что у стегоаналитика имеется в распоряжении некоторое число пустых и заполненных изображений, составляющих обучающее множество (следует из первого требования сценария: стегоаналитик может внедрить случайную информацию в пустые изображения). Далее стегоаналитик действует по следующему алгоритму:

1) извлечь признаки из изображений, составляющих обучающее множество;
2) обучить классификатор различать признаки пустых и заполненных изображений;
3) извлечь признаки очередного изображения из контрольного множества и с помощью обученного классификатора отнести его к классу пустых/заполненных.

1.2. Ансамблевый классификатор и его элементы

Предлагаемые в работе методы опираются на идею применения ансамблевых классификаторов к задачам стегоанализа [4]. Ансамблевые классификаторы называют «отличной альтернативой методу опорных векторов» из-за их хорошей производительности и конкурентоспособной эффективности [15]. Эти классификаторы, в частности, применялись победителями известного конкурса по стегоанализу BOSS competition [15]. Схема работы ансамблевого классификатора, как она описана в [4], следующая:

1) взять d признаков (таких, как SRM [1], SPAM [10], PSRM [16] и т. д.);
2) получить L случайно выбранных подмножеств из множества всех признаков, каждое из которых состоит из $d_{\text{sub}} < d$ признаков;
3) обучить L элементов ансамблевого классификатора на обучающем множестве различать пустые/заполненные изображения.

Пусть $N_{\text{votes}}(z)$ — число элементов ансамбля, голосующих за принадлежность изображения z классу пустых изображений:

$$N_{\text{votes}}(z) = \sum_{l=1}^{L} B_l(z).$$

Каждый элемент ансамбля работает следующим образом:

$$B_l(z) = \begin{cases} 1, & \text{элемент } l \text{ голосует за то, что } z \text{ — пустое,} \\ 0, & \text{элемент } l \text{ голосует за то, что } z \text{ — заполненное.} \end{cases}$$

Решение о принадлежности очередного тестового изображения к тому или иному классу принимается согласно следующему правилу:

$$\text{Ensemble-Rule}(L, N_{\text{votes}}) = \begin{cases} 1, & \text{если } N_{\text{votes}} > L/2, \\ 0, & \text{если } N_{\text{votes}} < L/2, \\ \text{random}\{0,1\} & \text{иначе.} \end{cases}$$

1.3. Формирование обучающего и контрольного множеств

Далее при проведении экспериментов обучающее и контрольное множества формируются на основе одной из наиболее известных баз изображений BOSSbase v1.01 [6, 7], которая часто используется специалистами по стеганографии и стегоанализу в качестве источника изображений. Данная база содержит 10000 чёрно-белых 8-битовых изображений размера 512×512 пикселей.

Обозначим обучающее и контрольное множества через $\mathcal{X}^p$ и $\mathcal{Y}^p$ соответственно, где p указывает на размер внедрения в битах на пиксель, и опишем процесс их формирования:

1) всё множество BOSSbase v1.01 разделено на два подмножества $\mathcal{X}_0$ и $\mathcal{Y}_0$, где $|\mathcal{X}_0| = 8000$ и $|\mathcal{Y}_0| = 2000$;
2) посредством случайного внедрения p б/п во все изображения из $\mathcal{X}_0$ и $\mathcal{Y}_0$ получены множества $\mathcal{X}_1^p$ и $\mathcal{Y}_1^p$ соответственно;
3) обучающее множество формируется как $\mathcal{X}^p = \mathcal{X}_0 \cup \mathcal{X}_1^p$;
4) контрольное множество формируется как $\mathcal{Y}^p = \mathcal{Y}_0 \cup \mathcal{Y}_1^p$.

Таким образом, $|\mathcal{X}^p| = 16000$, $|\mathcal{Y}^p| = 4000$ и в каждом множестве половина изображений пусты, половина заполнены. Далее индекс p, обозначающий размер внедрения, будем опускать (это не должно вызвать путаницы у читателя), и множества обозначаются через $\mathcal{X} = \mathcal{X}_0 \cup \mathcal{X}_1$ (обучающее) и $\mathcal{Y} = \mathcal{Y}_0 \cup \mathcal{Y}_1$ (контрольное).

1.4. Используемые в классификации признаки

Как известно, в методах машинного обучения применяются признаки, извлекаемые из объектов классификации. В настоящей работе в качестве таких признаков берутся SRM-признаки (Spatial Rich Model) [1], позволяющие создавать одни из наиболее эффективных методов стегоанализа. Их более новый вариант, PSRM-признаки (Projection Spatial Rich Model) [16], снижают ошибку обнаружения лишь незначительно, но при этом повышают сложность методов, существенно замедляя работу. Размерность пространства SRM-признаков составляет 34,671. Программа для извлечения этих признаков из изображений взята с сайта [17].

1.5. Элементы ансамбля

Существуют разные варианты выбора элементов ансамбля, но в экспериментах мы следуем рекомендациям [4] и используем линейный дискриминант Фишера (Fisher Linear Discriminant) [18] в силу его быстрого обучения и хороших результатов, которые показывают методы стегоанализа на его основе. Применяются два типа элементов ансамбля, которые обозначаются соответственно через

$$B_l,\ l = 1, \ldots, L \quad \text{и} \quad B'_m,\ m = 1, \ldots, M.$$

Таким образом, в первом случае их число равно L, а во втором — M. Каждому элементу приписывается по 800 случайно выбранных SRM-признаков.

1.6. Алгоритм внедрения информации

Эффективность предлагаемого подхода исследуется посредством его применения к обнаружению информации, внедрённой с помощью адаптивного метода Highly Undetectable Steganography (HUGO) [19]. На сегодняшний день этот метод считается наиболее трудно обнаружимым (см., например, результаты [16], где HUGO сравнивается с другими методами адаптивного внедрения, такими, как WOW [20] и UNIWARD [21]). HUGO базируется на ±1-стеганографии (LSB matching), но при его использовании места для внедрения выбираются не случайно, а вероятностно в зависимости от SPAM-признаков [19]. Такая модификация позволяет увеличить размер внедряемого сообщения приблизительно в 7 раз по сравнению с внедрением при помощи ±1-стеганографии с сохранением уровня стойкости (другими словами, при такой же ошибке обнаружения).

1.7. Лучшие результаты обнаружения метода HUGO

Для того чтобы оценить эффективность предварительной фильтрации, очевидно, необходимо сравнить ошибку обнаружения, вычисленную по всему множеству, с ошибкой обнаружения, вычисленной по подмножеству, полученному после фильтрации. При этом необходимо иметь реализацию ансамблевого классификатора, чтобы вычислять эти ошибки. Однако поскольку классификатор имеет достаточно много параметров, а в литературе они не всегда приводятся, то мы реализовали ансамблевый классификатор самостоятельно. Для проверки правильности реализации и достоверности вычисленных для него ошибок обнаружения мы сравнили ошибку обнаружения для этой реализации с ошибками обнаружения лучших реализаций [16]. Данные, приведённые в табл. 1, показывают, что значения ошибок обнаружения для нашей реализации согласуются с существующими данными. Далее при сравнении результатов предварительной фильтрации будем ориентироваться на ошибки обнаружения для нашей реализации (0,44, 0,37 и 0,141), поскольку ошибка обнаружения для «хорошего» подмножества вычисляется с помощью неё же.

Таблица 1

Лучшие результаты обнаружения метода HUGO (ансамблевый классификатор)

Размер внедрения, б/п	Ошибка обнаружения, вычисленная по всему контрольному множеству (BOSSbase v1.01)	
	Результаты из [16] при различных параметрах	Наша реализация, SRM-признаки, $L = 500$
0,05	—	0,44
0,10	0,3564–0,3757	0,37
0,20	0,2397–0,2701	—
0,40	0,1172–0,1383	0,141

Далее $\mathcal{Y}^{\text{good}}$ — подмножество «хороших» изображений, которые отбираются после предварительной фильтрации, а $P_{\text{E}}(\mathcal{Y}^{\text{good}})$ — ошибка обнаружения, вычисленная по данному подмножеству. В экспериментах сравниваются $P_{\text{E}}(\mathcal{Y}^{\text{good}})$ и $P_{\text{E}}(\mathcal{Y})$ со стратегический целью снизить $P_{\text{E}}(\mathcal{Y}^{\text{good}})$ и увеличить $|\mathcal{Y}^{\text{good}}|$ — размер $\mathcal{Y}^{\text{good}}$.

2. Описание и экспериментальное обоснование предлагаемого подхода

2.1. Базовая идея

Основной идеей, на которую опираются все методы, предлагаемые в работе, является естественное предположение о том, что если для некоторого изображения z значение $N_{\text{votes}}(z)$ близко к 0 или к L, то можно быть более уверенным в решении, чем в случае, когда это значение далеко от 0 и от L. Другими словами, если элементов ансамбля, проголосовавших за наличие информации, очень мало, то, скорее всего, её там действительно нет, а если таких голосов много, то, скорее всего, она там есть.

Данная идея непосредственно реализована в первом предлагаемом методе, который выбирает «хорошие» изображения, для которых

$$N_{\text{votes}}(z) \leqslant T^{\text{left}} \quad \text{или} \quad N_{\text{votes}}(z) \geqslant T^{\text{right}}$$

для некоторых заданных порогов T^{left} и T^{right}. Следующий метод, названный «простой классификацией», заключается в том, чтобы обучить некий дополнительный классификатор различать между «хорошими» и «плохими» изображениями и использовать его для выбора «хороших» изображений. Наконец, третий метод является комбинацией первых двух.

2.2. Метод 1: наивный

Идея, воплощённая в данном методе, заключается в определении двух порогов T^{left} и T^{right}, таких, что T^{left} близок к 0, а T_{right} близок к L (количеству элементов ансамбля), и разделении контрольного множества $\mathcal{Y}$ на «хорошее» и «плохое» подмножества согласно этим порогам следующим образом (алгоритм 1): $\mathcal{Y} = \mathcal{Y}^{\text{good}} \cup \mathcal{Y}^{\text{bad}}$, где

$$\mathcal{Y}^{\text{good}} = \{y \in \mathcal{Y} : N_{\text{votes}}(y) \leqslant T^{\text{left}} \text{ или } N_{\text{votes}}(y) \geqslant T^{\text{right}}\}, \quad \mathcal{Y}^{\text{bad}} = \mathcal{Y} \setminus \mathcal{Y}^{\text{good}}.$$

Алгоритм 1. Наивный метод фильтрации Naive-Method$(\mathcal{Z}, T^{\text{left}}, T^{\text{right}})$

Вход: $\mathcal{Z}$ — множество, из которого выделяются «хорошие» изображения, T^{left} и T^{right} — левый и правый пороги соответственно

Выход: $\mathcal{Z}^{\text{good}} \subseteq \mathcal{Z}$ — подмножество «хороших» изображений

1: Обучить элементы ансамбля $B_l, \ldots, B_L$ на подмножествах обучающей выборки $\mathcal{X}_0$ и $\mathcal{X}_1$ различать пустые/заполненные изображения

2: Для каждого изображения $z \in \mathcal{Z}$ вычислить число элементов ансамбля, проголосовавших за то, что изображение является пустым: $N_{\text{votes}}(z) = \sum_{l=1}^{L} B_l(z)$

3: Сформировать подмножество

$$\mathcal{Z}^{\text{good}} = \{z \in \mathcal{Z} : N_{\text{votes}}(z) \leqslant T^{\text{left}} \text{ или } N_{\text{votes}}(z) \geqslant T^{\text{right}}\}$$

При проведении экспериментов в качестве параметра $\mathcal{Z}$ в алгоритм 1 передавалось контрольное множество $\mathcal{Y}$. Рабочая гипотеза о том, что данная фильтрация позволит выделить подмножество $\mathcal{Y}^{\text{good}}$, такое, что $P_{\text{E}}(\mathcal{Y}^{\text{good}}) < P_{\text{E}}(\mathcal{Y})$, в целом оправдалась, однако по-настоящему впечатляющими результаты оказались только при анализе внедрения HUGO 0,40 б/п. Во-первых, множество «хороших» изображений оказалось достаточно большим (более 32 % от $\mathcal{Y}$). Во-вторых, ошибка обнаружения снизилась значительно (в зависимости от размера внедрения она лежит в промежутке

0,0016–0,0042), что приблизительно в 50 раз меньше, чем $P_{\mathrm{E}}(\mathcal{Y}) = 0{,}141$ (см. табл. 1). Для внедрения HUGO 0,05 и 0,10 б/п ошибка $P_{\mathrm{E}}(\mathcal{Y}^{\mathrm{good}})$ также меньше, чем $P_{\mathrm{E}}(\mathcal{Y})$, но различие не такое существенное и, кроме того, размеры подмножеств $P_{\mathrm{E}}(\mathcal{Y}^{\mathrm{good}})$ крайне малы.

Таблица 2

Наивный метод ($T^{\mathbf{left}} = 1$, $T^{\mathbf{right}} = L - 1$, % — доля (в %) «хороших» изображений)

	HUGO 0,05 б/п			HUGO 0,10 б/п			HUGO 0,40 б/п		
L	$\lvert\mathcal{Y}^{\mathrm{good}}\rvert$	%	$P_{\mathrm{E}}(\mathcal{Y}^{\mathrm{good}})$	$\lvert\mathcal{Y}^{\mathrm{good}}\rvert$	%	$P_{\mathrm{E}}(\mathcal{Y}^{\mathrm{good}})$	$\lvert\mathcal{Y}^{\mathrm{good}}\rvert$	%	$P_{\mathrm{E}}(\mathcal{Y}^{\mathrm{good}})$
100	50	1,25	0,260	271	6,76	0,140	1651	41,28	0,0042
200	31	0,76	0,258	176	4,40	0,125	1462	36,55	0,0041
300	23	0,56	0,304	137	3,43	0,124	1384	34,60	0,0022
400	16	0,40	0,313	117	2,93	0,120	1323	33,08	0,0023
500	14	0,35	0,286	106	2,65	0,123	1285	32,13	0,0016

2.3. Метод 2: простая классификация

С расчётом увеличить размер подмножества $\mathcal{Y}^{\mathrm{good}}$, особенно для внедрения HUGO 0,05 и 0,10 б/п, мы решили обучить дополнительный ансамблевый классификатор, который мог бы различать «хорошие» и «плохие» изображения (алгоритм 2). Идея метода заключается в том, чтобы сначала разделить обучающее множество на «хорошее» и «плохое» подмножества с помощью наивного метода (алгоритм 1), а затем обучить дополнительный классификатор на этом разбиении.

Алгоритм 2. Метод простой классификации Simple-Classification($\mathcal{Z}, T^{\mathrm{left}}, T^{\mathrm{right}}$)

Вход: $\mathcal{Z}$ — множество, из которого выделяются «хорошие» изображения, T^{left} и T^{right} — левый и правый пороги соответственно

Выход: $\mathcal{Z}^{\mathrm{good}} \subseteq \mathcal{Z}$ — подмножество «хороших» изображений

1: Получить подмножество «хороших» изображений из обучающего множества с помощью наивного метода $\mathcal{Z}^{\mathrm{good}} :=$ Naive-Method($\mathcal{Z}, T^{\mathrm{left}}, T^{\mathrm{right}}$)
2: Получить подмножество «плохих» изображений $\mathcal{Z}^{\mathrm{bad}} := \mathcal{Z} \setminus \mathcal{Z}^{\mathrm{good}}$
3: Обучить элементы дополнительного ансамбля $B'_1, \ldots, B'_M$ различать «хорошие»/«плохие» изображения на разбиении $\mathcal{Z}^{\mathrm{good}}$ (класс 0) и $\mathcal{Z}^{\mathrm{bad}}$ (класс 1)
4: Получить подмножество $\mathcal{Z}^{\mathrm{good}} = \{z \in \mathcal{Z} : \text{Ensemble-Rule}(M, N'_{\mathrm{votes}}(z)) = 0\}$

Как показали эксперименты (табл. 3), простая классификация действительно позволила значительно увеличить размер подмножества $\mathcal{Y}^{\mathrm{good}}$ по сравнению с наивным методом (табл. 2), хотя ошибка обнаружения $P_{\mathrm{E}}(\mathcal{Y}^{\mathrm{good}})$ уменьшилась незначительно по сравнению с результатами, приведёнными в табл. 1. Более того, для внедрения HUGO 0,40 б/п она даже возросла.

2.4. Метод 3: комбинированная классификация

Для получения более эффективного метода предварительной фильтрации, который позволил бы снизить ошибку обнаружения и увеличить размер множества «хороших» изображений (как минимум для HUGO 0,05 б/п и HUGO 0,10 б/п, поскольку для HUGO 0,40 б/п наивный метод уже показал впечатляющие результаты), создана комбинация двух предложенных выше методов. Идея комбинированного метода заключается в том, чтобы выделить «хорошие» подмножества наивным методом и методом

Таблица 3

Простая классификация (% — доля «хороших» изображений)

T^{left}	T^{right}	HUGO 0,05 б/п			HUGO 0,10 б/п			HUGO 0,40 б/п		
		$\lvert\mathcal{Y}^{\text{good}}\rvert$	%	$P_{\text{E}}(\mathcal{Y}^{\text{good}})$	$\lvert\mathcal{Y}^{\text{good}}\rvert$	%	$P_{\text{E}}(\mathcal{Y}^{\text{good}})$	$\lvert\mathcal{Y}^{\text{good}}\rvert$	%	$P_{\text{E}}(\mathcal{Y}^{\text{good}})$
1	499	292	7	0,353	1198	30	0,244	1903	48	0,0189
2	499	280	7	0,346	1139	28	0,227	2022	51	0,0218
3	499	455	11	0,365	1158	29	0,225	2061	52	0,0213
1	498	232	6	0,332	1510	38	0,273	1911	48	0,0167
2	498	337	8	0,359	1189	30	0,230	2132	53	0,0225
3	498	528	13	0,371	1302	33	0,247	2009	50	0,0184
1	497	284	7	0,357	1171	29	0,243	2045	51	0,0220
2	497	340	9	0,362	1204	30	0,236	2091	52	0,0210
3	497	347	9	0,378	1182	30	0,228	2048	51	0,0215

простой классификации, а затем взять их пересечение. Согласно рабочей гипотезе, при удачном подборе порогов T_1^{left} и T_1^{right} для наивного метода и T_2^{left} и T_2^{right} для метода простой классификации окажется возможным выбрать достаточно большое «хорошее» подмножество $\mathcal{Y}^{\text{good}}$, для которого ошибка обнаружения $P_{\text{E}}(\mathcal{Y}^{\text{good}})$ будет заметно меньше, чем по всему контрольному множеству. Описание метода приведено в алгоритме 3.

Алгоритм 3. Метод комбинированной классификации Combined-Classification$(\mathcal{Z}, T_1^{\text{left}}, T_1^{\text{right}}, T_2^{\text{left}}, T_2^{\text{right}})$

Вход: $\mathcal{Z}$ — множество, из которого выделяются «хорошие» изображения, T_1^{left} и T_1^{right} — пороги для наивного метода, T_2^{left} и T_2^{right} — пороги для простой классификации

Выход: $\mathcal{Z}^{\text{good}} \subseteq \mathcal{Z}$ — подмножество «хороших» изображений

1: $\mathcal{Z}_1^{\text{good}} = $ Naive-Method$(\mathcal{Z}, T_1^{\text{left}}, T_1^{\text{right}})$
2: $\mathcal{Z}_2^{\text{good}} = $ Simple-Classification$(\mathcal{Z}, T_2^{\text{left}}, T_2^{\text{right}})$
3: $\mathcal{Z}^{\text{good}} = \mathcal{Z}_1^{\text{good}} \cap \mathcal{Z}_2^{\text{good}}$

Схема экспериментов следующая: фиксировалась максимально допустимая ошибка P_{E}^* и подбирались пороги, обеспечивающие максимальный размер подмножества «хороших» изображений $\mathcal{Y}^{\text{good}}(T_1^{\text{left}}, T_1^{\text{right}}, T_2^{\text{left}}, T_2^{\text{right}})$ при условии, что $P_{\text{E}}(\mathcal{Y}^{\text{good}}) \leqslant P_{\text{E}}^*$.

Более формально, $(T_2^{\text{left}}(P_{\text{E}}^*), T_2^{\text{right}}(P_{\text{E}}^*)) = \arg\max_{t^{\text{left}}, t^{\text{right}}} |\mathcal{Y}^{\text{good}}(T_l^{\text{left}}, T_l^{\text{right}}, t^{\text{left}}, t^{\text{right}})|$ при условии, что $P_{\text{E}}(\mathcal{Y}^{\text{good}}(T_l^{\text{left}}, T_l^{\text{right}}, t^{\text{left}}, t^{\text{right}})) \leqslant P_{\text{E}}^*$.

В табл. 4–7 представлены результаты экспериментов по комбинированной классификации. Параметры: табл. 4 — HUGO 0,1 б/п, $L = 500$, $M = 1$; табл. 5 — HUGO 0,05 б/п, $L = 500$, $M = 11$, $T_1^{\text{left}} = 10$, $T_1^{\text{right}} = 490$; табл. 6 — HUGO 0,4 %, $L = 500$, $M = 11$, $T_1^{\text{left}} = 20$, $T_1^{\text{right}} = 480$; табл. 7 — HUGO 0,1 б/п, $L = 500$, $M = 11$.

В табл. 7 показано, что из всего множества BOSSbase можно выбрать 5 % изображений, для которых ошибка обнаружения стеганографии HUGO 0,1 б/п не превосходит 0,05, в то время как ошибка по всему множеству составляет 0,37 (см. табл. 1), что нельзя считать достоверным обнаружением, поскольку значение близко к 0,5 (случайному угадыванию).

Таблица 4

	$T_1^{\text{left}}=1,\ T_1^{\text{right}}=499$				$T_1^{\text{left}}=2,\ T_1^{\text{right}}=498$			
P_{E}^*	$\lvert\mathcal{Y}^{\text{good}}\rvert$	%	T_2^{left}	T_2^{right}	$\lvert\mathcal{Y}^{\text{good}}\rvert$	%	T_2^{left}	T_2^{right}
0,04	187	5	1	489	202	12	2	490
0,05	230	6	2	485	251	12	2	481
0,06	252	6	4	485	303	12	17	490
0,07	346	8	19	483	391	12	27	481
0,08	401	10	33	489	463	12	30	464

Таблица 5

P_{E}^*	$\lvert\mathcal{Y}^{\text{good}}\rvert$	%	T_2^{left}	T_2^{right}
0,15	21	0,5	0	487
0,18	28	0,7	3	487
0,21	58	1,5	20	486
0,24	92	2,3	42	470

Таблица 6

P_{E}^*	$\lvert\mathcal{Y}^{\text{good}}\rvert$	%	T_2^{left}	T_2^{right}
0,00000	1481	37	1	492
0,00125	1655	41	9	492
0,00225	1680	42	9	490
0,00325	1853	46	99	492

Таблица 7

	$T_1^{\text{left}}=1,\ T_1^{\text{right}}=499$				$T_1^{\text{left}}=2,\ T_1^{\text{right}}=498$				$T_1^{\text{left}}=10,\ T_1^{\text{right}}=490$			
P_{E}^*	$\lvert\mathcal{Y}^{\text{good}}\rvert$	%	T_2^{left}	T_2^{right}	$\lvert\mathcal{Y}^{\text{good}}\rvert$	%	T_2^{left}	T_2^{right}	$\lvert\mathcal{Y}^{\text{good}}\rvert$	%	T_2^{left}	T_2^{right}
0,01	0	0	—	—	0	0	—	—	0	0	—	—
0,02	157	3,9	1	471	0	0	—	—	0	0	—	—
0,03	181	4,5	2	464	175	4,4	1	485	0	0	—	—
0,04	191	4,8	2	455	227	5,7	1	464	203	5,1	2	490
0,05	209	5,2	3	438	255	6,4	2	455	284	7,1	2	470
0,06	267	6,7	27	455	336	8,4	28	464	334	8,4	27	490
0,07	293	7,3	45	464	388	9,7	45	455	460	11,5	27	455
0,08	354	8,9	90	438	456	11,4	93	442	518	13,0	33	434
0,09	378	9,5	90	403	482	12,1	96	418	567	14,2	33	403
0,10	386	9,7	98	403	499	12,5	98	401	626	15,7	93	453
0,11	388	9,7	98	401	499	12,5	98	401	710	17,6	95	403

2.5. Предварительная фильтрация «на лету»

Описания предлагаемых методов предварительной фильтрации, приведённые выше, выполнены в терминах множеств и подмножеств, однако эти методы очевидным образом могут быть применены и к отдельным изображениям. Для этого вместо формирования подмножества $\mathcal{Z}^{\text{good}}$ можно тестировать каждое очередное изображение, что даст возможность проводить предварительную фильтрацию «на лету».

3. Возможные применения предварительной фильтрации и пути дальнейших исследований

Предварительная фильтрация может быть использована для повышения практической значимости слабых методов стегоанализа, которые дают ошибку обнаружения, близкую к 0,5 (вероятности случайного угадывания). Предварительная фильтрация может позволить отобрать некоторое подмножество «хороших» изображений, на котором ошибка обнаружения будет меньше. Например, подобные результаты получены в экспериментах с HUGO 0,05 б/п, когда ошибка по всему контрольному множеству равна 0,37 (см. табл. 1), что вряд ли может считаться достоверным обнаружением, а предварительная фильтрация позволила отобрать подмножество (хотя и не слишком большое) со значительно меньшей ошибкой (см. табл. 5).

Ещё одним потенциальным применением предварительной фильтрации может стать выбор наиболее достоверного метода обнаружения для заданного изображения или множества изображений. Например, если в распоряжении стегоаналитика имеется несколько методов обнаружения, то он может разбить контрольное множество на несколько подмножеств, каждое из которых обеспечивает низкую ошибку для определённого метода, и проводить стегоанализ этих подмножеств соответствующим методом. Итоговая ошибка обнаружения у такой схемы может быть ниже, чем у различных методов в отдельности.

На эффективность обнаружения информации влияют как минимум два фактора: метод обнаружения и специфические свойства изображения, поэтому, если один метод работает лучше на одном множестве изображений, то это не гарантирует того, что он будет работать лучше и на другом множестве. Таким образом, сравнивая несколько различных методов обнаружения скрытой информации, разумно проверять их на нескольких множествах с различными свойствами. Предварительная фильтрация может стать подходом, который позволит выделять такие множества, причём делить можно не только на «хорошие» и «плохие», но и более тонко, скажем, на «очень хорошие», «хорошие», «плохие» и «очень плохие». Как показано в экспериментах, с помощью выбора параметров можно задать требуемый размер этих множеств.

Если после предварительной фильтрации доля «хороших» изображений достаточно велика, то можно составить схему, которая будет иметь более высокую производительность, чем некий высокоточный, но медленный метод. Например, если один метод обнаружения работает медленно (как метод опорных векторов), но является эффективным, а другой метод работает быстрее (как ансамблевый классификатор), то с помощью быстрого метода можно осуществить предварительную фильтрацию, а затем медленный метод будет обрабатывать только «хорошие» изображения.

Вычисление ошибки обнаружения по подмножеству аналогично разработке атак на криптосистемы в предположении использования слабых ключей [11, 12]. При этом размер множества слабых ключей является дополнительной характеристикой атаки: чем больше множество, тем эффективнее атака. Таким же образом размер подмножества «хороших» изображений может считаться дополнительной характеристикой метода обнаружения.

Заключение

В работе предложен новый подход к стегоанализу, подразумевающий предварительную фильтрацию изображений, подлежащих проверке на наличие в них скрытой информации. Данный подход заключается в том, что перед финальным этапом стегоанализа производится отбор изображений с целью выделить подмножество тех из них, которые обеспечат как можно меньшую ошибку обнаружения (во всяком случае, меньше ошибки, вычисленной по всему контрольному множеству). Предварительная фильтрация может быть реализована различными способами; здесь предложены три возможности: наивный метод, метод простой классификации и метод комбинированной классификации (комбинация первых двух).

Согласно проведённым экспериментам, предварительная фильтрация довольно чувствительна к выбору параметров (порогов), поэтому результаты представлены в виде таблиц, отражающих зависимости ошибки обнаружения от этих параметров. По нашему мнению, наиболее интересными результатами являются следующие. Эксперименты показали, что предварительная фильтрация позволяет выбрать порядка 35 % изображений из BOSSbase v1.01, для которых метод адаптивной стеганографии HUGO

0,4 б/п обнаруживается с ошибкой менее 0,003, в то время как ошибка, вычисленная по всему множеству, составляет 0,141. Показано также, что из всего множества можно выбрать порядка 5 % изображений, для которых HUGO 0,1 б/п определяется с ошибкой менее чем 0,05, тогда как ошибка по всему множеству составляет 0,37.

В работе описаны эти и другие потенциальные применения предварительной фильтрации, такие, как расширенное определение точности методов обнаружения скрытой информации, возможность выбирать размер «хорошего» подмножества за счёт настройки параметров, возможность деления множества изображений на подмножества с разными свойствами.

ЛИТЕРАТУРА

1. *Fridrich J.* Rich models for steganalysis of digital images // IEEE Trans. Information Forensics and Security. 2012. V. 7. No. 3. P. 868–882.
2. *Fridrich J., Kodovsky J., Holub V., and Goljan M.* Steganalysis of content-adaptive steganography in spatial domain // Proc. 13th Information Hiding Workshop. LNCS. 2011. V. 6958. P. 102–117.
3. *Ker A., Bas P., Bohme R., et al.* Moving steganography and steganalysis from the laboratory into the real world // Proc. 1st ACM Workshop on Information Hiding and Multimedia Security. N. Y., USA, 2013. P. 45–58.
4. *Kodovsky J., Fridrich J., and Holub V.* Ensemble classifiers for steganalysis of digital media // IEEE Trans. Information Forensics and Security. 2011. V. 7. No. 2. P. 434–444.
5. *Monarev V. and Pestunov A.* A known-key scenario for steganalysis and a highly accurate detector within it // Proc. 10th IEEE Intern. Conf. Intelligent Information Hiding and Multimedia Signal Processing. Kitakyushu, 2014. P. 175–178.
6. `www.agents.cz/boss` — Break our steganographic system. 2015.
7. *Bas P., Filler T., and Pevny T.* Break our steganographic system — the ins and outs of organizing BOSS // Proc. 13th Information Hiding Workshop. LNCS. 2011. V. 6958. P. 59–70.
8. `bows2.ec-lille.fr` — Break our watermarking system, second edition. 2015.
9. `photogallery.nrcs.usda.gov/res/sites/photogallery` — NRCS photo gallery. 2015.
10. *Pevny T., Bas P., and Fridrich J.* Steganalysis by subtractive pixel adjacency matrix // IEEE Trans. Information Forensics and Security. 2010. V. 5. No. 2. P. 215–224.
11. *Biryukov A., Nakahara J., Preneel B., and Vanderwalle J.* New weak-key classes of IDEA // LNCS. 2002. V. 2513. P. 315–326.
12. *Kara O. and Manap C.* A new class of weak keys for Blowfish // FSE'2007. LNCS. 2007. V. 4593. P. 167–180.
13. *Pevny T.* Detecting messages of unknown length // Proc. 8th Intern. Conf. Media Watermarking, Security and Forensics. 2011. P. 1–12.
14. *Monarev V. and Pestunov A.* A new compression-based method for estimating LSB replacement rate in color and grayscale images // Proc. 7th IEEE Intern. Conf. Intelligent Information Hiding and Multimedia Signal Processing. Dalian, 2011. P. 57–60.
15. *Fridrich J., Kodovsky J., Holub V., and Goljan M.* Breaking HUGO — the process discovery // Proc. 13th Information Hiding Workshop. LNCS. 2011. V. 6958. P. 102–117.
16. *Holub V. and Fridrich J.* Random projections of residuals for digital image steganalysis // IEEE Trans. Information Forensics and Security. 2013. V. 8. No. 12. P. 1996–2006.
17. `dde.binghamton.edu/download/feature_extractors` — Feature extractors for steganalysis. 2015.
18. *Duda R., Hart P., and Stork D.* Pattern Classification. 2nd ed. N. Y.: John Wiley & Sons Inc., 2001.

19. *Pevny T., Filler T., and Bas P.* Using high-dimensional image models to perform highly undetectable steganography // Proc. 12th Information Hiding Workshop. LNCS. 2010. V. 6387. P. 161–177.
20. *Holub V. and Fridrich J.* Designing steganographic distortion using directional filters // Proc. 4th IEEE Intern. Workshop on Information Forensics and Security. Tenerife, 2012. P. 234–239.
21. *Holub V. and Fridrich J.* Digital image steganography using universal distortion // Proc. 1th ACM Workshop on Information Hiding and Multimedia Security. N. Y., USA, 2013. P. 59–68.

REFERENCES

1. *Fridrich J.* Rich models for steganalysis of digital images. IEEE Trans. Information Forensics and Security, 2012, vol. 7, no. 3, pp. 868–882.
2. *Fridrich J., Kodovsky J., Holub V., and Goljan M.* Steganalysis of content-adaptive steganography in spatial domain. Proc. 13th Information Hiding Workshop, LNCS, 2011, vol. 6958, pp. 102–117.
3. *Ker A., Bas P., Bohme R., et al.* Moving steganography and steganalysis from the laboratory into the real world. Proc. 1st ACM Workshop on Information Hiding and Multimedia Security, N. Y., USA, 2013, pp. 45–58.
4. *Kodovsky J., Fridrich J., and Holub V.* Ensemble classifiers for steganalysis of digital media. IEEE Trans. Information Forensics and Security, 2011, vol. 7, no. 2, pp. 434–444.
5. *Monarev V. and Pestunov A.* A known-key scenario for steganalysis and a highly accurate detector within it. Proc. 10th IEEE Intern. Conf. Intelligent Information Hiding and Multimedia Signal Processing, Kitakyushu, 2014, pp. 175–178.
6. `www.agents.cz/boss` — Break our steganographic system. 2015.
7. *Bas P., Filler T., and Pevny T.* Break our steganographic system — the ins and outs of organizing BOSS. Proc. 13th Information Hiding Workshop, LNCS, 2011, vol. 6958, pp. 59–70.
8. `bows2.ec-lille.fr` — Break our watermarking system, second edition. 2015.
9. `photogallery.nrcs.usda.gov/res/sites/photogallery` — NRCS photo gallery. 2015.
10. *Pevny T., Bas P., and Fridrich J.* Steganalysis by subtractive pixel adjacency matrix. IEEE Trans. Information Forensics and Security, 2010, vol. 5, no. 2, pp. 215–224.
11. *Biryukov A., Nakahara J., Preneel B., and Vanderwalle J.* New weak-key classes of IDEA. LNCS, 2002, vol. 2513, pp. 315–326.
12. *Kara O. and Manap C.* A new class of weak keys for Blowfish FSE'2007. LNCS. 2007, vol. 4593, pp. 167–180.
13. *Pevny T.* Detecting messages of unknown length. Proc. 8th Intern. Conf. Media Watermarking, Security and Forensics, 2011, pp. 1–12.
14. *Monarev V. and Pestunov A.* A new compression-based method for estimating LSB replacement rate in color and grayscale images. Proc. 7th IEEE Intern. Conf. Intelligent Information Hiding and Multimedia Signal Processing, Dalian, 2011, pp. 57–60.
15. *Fridrich J., Kodovsky J., Holub V., and Goljan M.* Breaking HUGO — the process discovery. Proc. 13th Information Hiding Workshop, LNCS, 2011, vol. 6958, pp. 102–117.
16. *Holub V. and Fridrich J.* Random projections of residuals for digital image steganalysis. IEEE Trans. Information Forensics and Security, 2013, vol. 8, no. 12, pp. 1996–2006.
17. `dde.binghamton.edu/download/feature_extractors` — Feature extractors for steganalysis. 2015.
18. *Duda R., Hart P., and Stork D.* Pattern Classification. 2nd ed. N. Y., John Wiley & Sons Inc., 2001.

19. *Pevny T., Filler T., and Bas P.* Using high-dimensional image models to perform highly undetectable steganography. Proc. 12th Information Hiding Workshop, LNCS, 2010, vol. 6387, pp. 161–177.
20. *Holub V. and Fridrich J.* Designing steganographic distortion using directional filters. Proc. 4th IEEE Intern. Workshop on Information Forensics and Security, Tenerife, 2012, pp. 234–239.
21. *Holub V. and Fridrich J.* Digital image steganography using universal distortion. Proc. 1th ACM Workshop on Information Hiding and Multimedia Security, N. Y., USA, 2013, pp. 59–68.

ПРИКЛАДНАЯ ТЕОРИЯ ГРАФОВ

УДК 519.172

ПРОВЕРКА ПЛАНАРНОСТИ И ПОСТРОЕНИЕ ТОПОЛОГИЧЕСКОГО РИСУНКА ПЛОСКОГО ГРАФА (ПОИСКОМ В ГЛУБИНУ)

С. В. Курапов, М. В. Давидовский

Запорожский национальный университет, г. Запорожье, Украина

Рассматривается алгоритм проверки графа на планарность с одновременным построением математических структур для описания топологического рисунка плоского графа. Такими математическими структурами являются изометрические циклы и вращение вершин графа. Показано, что система изометрических циклов графа индуцирует вращение вершин для описания топологического рисунка плоского графа. В отличие от классических алгоритмов проверки планарности, например алгоритма Хопкрофта — Тарьяна, полученный в результате работы алгоритма топологический рисунок используется для визуализации плоского графа. Вычислительная сложность алгоритма определяется как $O(m^2)$, где m — количество рёбер графа.

Ключевые слова: *граф, планарность, визуализация графа, топологический рисунок графа, алгоритмы на графах, вращение вершин, изометрические циклы.*

DOI 10.17223/20710410/32/7

PLANARITY TESTING AND CONSTRUCTING THE TOPOLOGICAL DRAWING OF A PLANE GRAPH (DFS)

S. V. Kurapov, M. V. Davidovsky

Zaporizhzhya National University, Zaporizhzhya, Ukraine

E-mail: lilili5050@rambler.ru, m.davidovsky@gmail.com

In this article we present a new graph planarity testing algorithm along with the construction of mathematical framework used for representing topological drawings of plane graphs. This mathematical framework is based on the notions of graph isometric cycles and rotation of graph vertices. It is shown that the system of isometric cycles of a graph induces the rotation of its vertices for representing topological drawing of the plane graph. In contrast to the classical planarity testing algorithms, e. g. the Hopcroft — Tarjan algorithm, the topological drawing obtained as a result of the proposed algorithm execution is used subsequently for the visualization of the planar graph. Computational complexity of the proposed algorithm is estimated by $O(m^2)$, where m is the number of edges in the graph.

Keywords: *graph, planarity, graph visualization, topological graph drawing, graph algorithms, vertices rotation, isometric cycles.*

Введение

Для решения прикладных задач, например при создании систем автоматизации проектирования плоских конструктивов [1, 2], необходимо иметь математический аппарат, описывающий рисунок плоского графа для визуализации его изображения [3–6]. Покажем, что такими математическими структурами являются изометрические циклы графа [7, 8] и вращение вершин графа [9]. Будем рассматривать простые несепарабельные неориентированные графы.

Определение 1. *Несепарабельным графом* будем называть связный неориентированный граф без петель и кратных рёбер, без мостов и точек сочленения, без вершин с локальной степенью меньшей или равной двум.

Пусть $G = (X, U)$ — несепарабельный граф с пронумерованным множеством рёбер $U = \{u_1, u_2, \ldots, u_m\}$ и вершин $X = \{x_1, x_2, \ldots, x_n\}$. Обычно граф G представляется матрицей инциденций или матрицей смежностей. Графически граф может быть представлен диаграммой, в которой вершина изображена точкой или кружком, а ребро — отрезком линии, соединяющим вершины [10–12]. В случае планарного графа всегда имеется возможность проведения соединений (рёбер графа) без пересечений, причём такое представление не зависит от расположения вершин и характера проведения соединений. Это представление планарного графа называется плоским изображением графа (рис. 1).

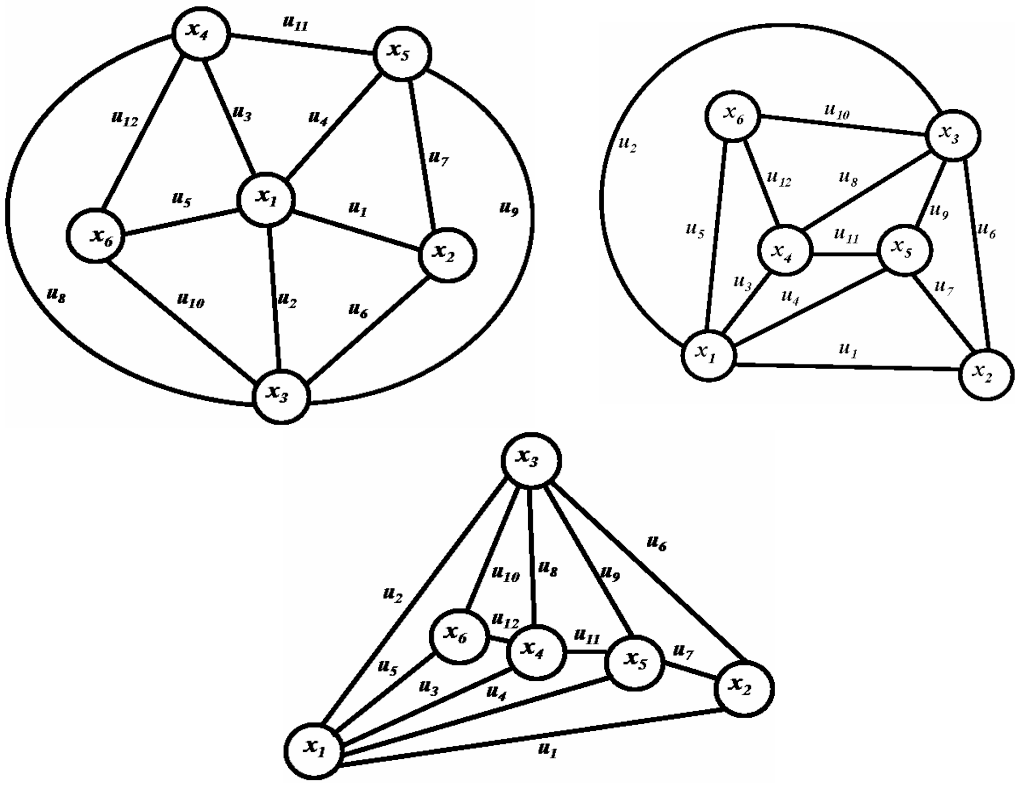

Рис. 1. Различные диаграммы графа G

1. Топологический рисунок графа

Существуют структуры, которые являются общими для любого плоского изображения графа. Рассмотрим множество простых циклов, являющихся границами граней плоского изображения. Как пример рассмотрим граф G на рис. 1. Запишем множество

граничных циклов в виде элементов пространства суграфов:

$$c_1 = \{u_2, u_5, u_{10}\}; \quad c_2 = \{u_3, u_5, u_{12}\}; \quad c_3 = \{u_8, u_{10}, u_{12}\}; \quad c_4 = \{u_8, u_9, u_{11}\};$$
$$c_5 = \{u_6, u_7, u_9\}; \quad c_6 = \{u_3, u_4, u_{11}\}; \quad c_7 = \{u_1, u_4, u_7\}; \quad c_8 = \{u_1, u_2, u_6\}.$$

Цикломатическое число определяет количество независимых циклов графа $\nu(G) = m - n + 1$. Кольцевая сумма независимых циклов определяет обод.

Если задать направление обхода рёбер в циклах с соблюдением условия планарности Маклейна [13], то можно записать циклы как кортежи вершин (рис. 2):

$$c_1 = \{u_2, u_5, u_{10}\} \to \langle x_1, x_3, x_6\rangle; \quad c_2 = \{u_3, u_5, u_{12}\} \to \langle x_1, x_6, x_4\rangle;$$
$$c_3 = \{u_8, u_{10}, u_{12}\} \to \langle x_4, x_6, x_3\rangle; \quad c_4 = \{u_8, u_9, u_{11}\} \to \langle x_4, x_3, x_5\rangle;$$
$$c_5 = \{u_6, u_7, u_9\} \to \langle x_3, x_2, x_5\rangle; \quad c_6 = \{u_3, u_4, u_{11}\} \to \langle x_4, x_5, x_1\rangle;$$
$$c_7 = \{u_1, u_4, u_7\} \to \langle x_1, x_5, x_2\rangle; \quad c_8 = \{u_1, u_2, u_6\} \to \langle x_1, x_2, x_3\rangle.$$

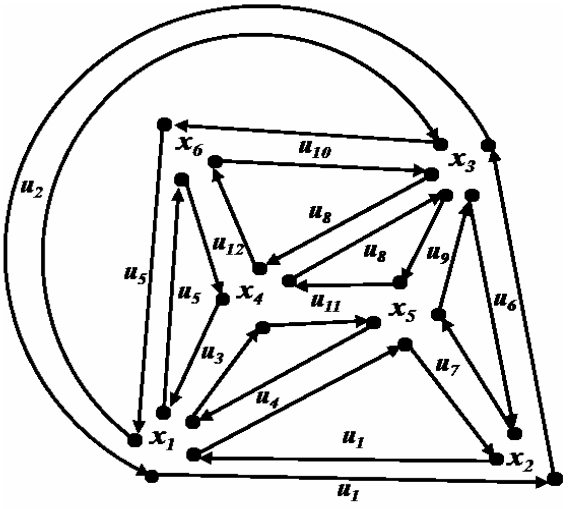

Рис. 2. Задание направления обхода рёбер в циклах

С другой стороны, заданное подмножество циклов с направлением обхода рёбер порождает (индуцирует) определённый циклический порядок расположения смежных вершин для каждой вершины (рис. 3).

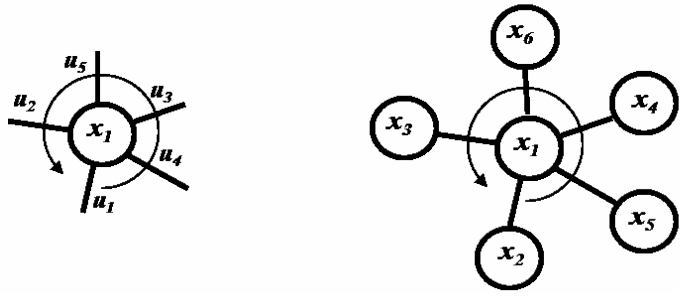

Рис. 3. Вращение вершины x_1

Определение 2. Для данного графа G *вращение вершины* A графа G — это ориентированный циклический порядок (или циклическая перестановка) всех рёбер, инцидентных вершине A.

Вращение графа можно описывать и представлять следующим образом. Выпишем циклическую перестановку соседей для каждой вершины x_i. Эта перестановка порождается вращением вершины x_i, которое является циклической перестановкой рёбер, инцидентных вершине x_i.

Вращение вершины x_i обозначим $\hbar_i$. Вращение всех вершин описывается диаграммой вращения вершин $\hbar$. В нашем случае диаграмма имеет следующий вид:

$$\begin{array}{lccccc}
\hbar_1: & x_2 & x_5 & x_4 & x_6 & x_3 \\
\hbar_2: & x_3 & x_5 & x_1 & & \\
\hbar_3: & x_1 & x_6 & x_4 & x_5 & x_2 \\
\hbar_4: & x_6 & x_1 & x_5 & x_3 & \\
\hbar_5: & x_4 & x_1 & x_2 & x_3 & \\
\hbar_6: & x_1 & x_4 & x_3 & &
\end{array}$$

Пусть x_1 — вершина, инцидентная ребру u_1 в графе G с вращением $\hbar$. Построим в графе G замкнутый маршрут

$$x_1, u_1, x_2, u_2, x_3, u_3, \ldots,$$

где вершина x_2 — второй конец ребра u_1, а ребро u_2 следует за ребром u_1 во вращении вершины x_2, определяемом вращением $\hbar$. Затем определяется x_3 как вершина, инцидентная ребру u_2 и не равная x_2. После этого в качестве u_3 выбирается ребро, следующее за ребром u_2 во вращении вершины x_3, и т. д. Закончим процесс в точности перед тем моментом, когда должна повториться пара x_1, u_1. Она должна повториться, ибо граф G конечный, а процесс однозначно определён и в обратном направлении, а именно: если часть $x_{t-1}, u_t, x_t, \ldots$ известна, то ребро u_{t-1} определяется вращением вокруг вершины x_{t-1}. Назовём такой замкнутый маршрут циклом, порождённым вершиной x_1 и ребром u_1 и индуцированным вращением $\hbar$.

Тем самым вращение вершин $\hbar$ индуцирует (порождает) простые циклы. В свою очередь, система независимых циклов и обод индуцируют вращение вершин $\hbar$.

Введём следующие понятия, связанные с метрикой графа.

Определение 3 [8]. *Изометрический подграф* — подграф G' графа G, у которого все расстояния внутри G' те же самые, что и в G.

Определение 4. *Изометрическим циклом* в графе называется простой цикл, для которого кратчайший путь между любыми двумя его вершинами состоит из рёбер этого цикла. *Изометрический цикл* — частный случай изометрического подграфа.

В таком цикле между любыми двумя его несмежными вершинами в графе G не существует маршрутов меньшей длины, чем маршруты, принадлежащие данному циклу.

Предположим, что планарный граф G вложен в евклидову плоскость, причем выполняется такое свойство: *граница любой внутренней грани — изометрический цикл графа* G. Теперь можно дать определение топологического рисунка графа.

Определение 5. *Топологическим рисунком* планарного графа назовём одинаково направленное вращение всех его вершин $\hbar$, индуцирующее множество изометрических циклов графа, которое удовлетворяет критерию планарности Маклейна [6].

Один из алгоритмов проверки графа на планарность разработан в 1970 г. Дж. Хопкрофтом и Р. Тарьяном [14]. Авторы предложили алгоритм, требующий $O(n \log n)$ единиц времени, который они, в конечном счёте, улучшили до $O(n)$, где n — количество вершин графа. Алгоритм устанавливает планарность графа, но его применение вызывает определённые затруднения при дальнейшем решении задачи визуализации рисунка графа.

В данной работе представлен алгоритм проверки графа на планарность с одновременным получением топологического рисунка графа. Определим необходимый понятийный аппарат и опишем алгоритм проверки графа на планарность, не вдаваясь в детальное описание вспомогательных алгоритмов.

2. Выделение DFS-дерева графа

Описание будем проводить на примере графа G, представленного на рис. 4. Алгоритмом поиска в глубину выбираем дерево графа (рис. 5) и пронумеровываем вершины (рис. 6).

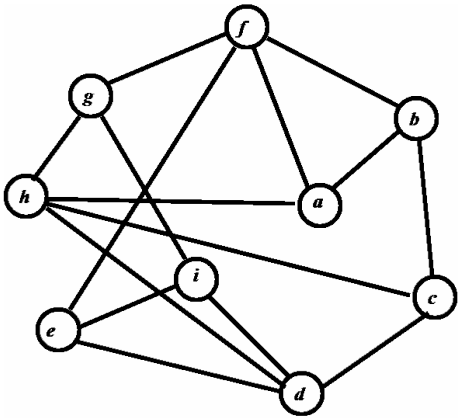

Рис. 4. Граф G

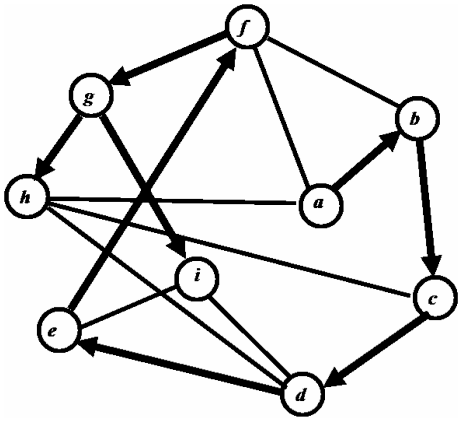

Рис. 5. DFS-дерево графа G

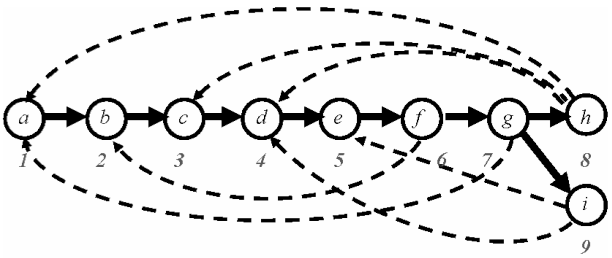

Рис. 6. Нумерация вершин графа G

Ориентированное остовное дерево, которое получается в результате поиска в глубину на простом неориентированном графе, будем называть *DFS-деревом* [15]. В результате построения такого DFS-дерева образуются обратные рёбра, которые на рис. 6 изображены пунктирными линиями. Они характерны тем, что номер вершины истока всегда больше номера вершины стока.

3. Построение опорного цикла и путей

Предлагаемый метод проверки графа на планарность основан на простом свойстве проверки пересечения обратных хорд для выделенного дерева графа. Используя

фундаментальную матрицу циклов, выберем самый длинный цикл, образованный ветвями дерева и одной хордой. Пусть это будет цикл $\langle a, b, c, d, e, f, g, h, a\rangle$. Данный цикл будем называть *опорным*. Опорный цикл состоит из ориентированных ветвей дерева $\langle a, b\rangle, \langle b, c\rangle, \langle c, d\rangle, \langle d, e\rangle, \langle e, f\rangle, \langle f, g\rangle, \langle g, h\rangle$ и ориентированного обратного ребра $\langle h, a\rangle$. Формируем X_1 — подмножество вершин, включённых в опорный цикл.

Определение 6. *Обратным ребром* называется ориентированный маршрут, состоящий из одной хорды.

Определение 7. *Обратным путём* называется ориентированный маршрут, состоящий из последовательно расположенных ветвей дерева и одной и только одной хорды; концевые вершины такого маршрута принадлежат опорному циклу.

Очевидно, что обратные рёбра являются частным случаем обратных путей. Множество обратных путей и обратных рёбер будем называть *множеством обратных маршрутов*.

Формируем маршруты, состоящие из обратных рёбер, вершины которых принадлежат опорному циклу. Это следующие обратные рёбра: $\langle g, a\rangle, \langle f, b\rangle, \langle h, c\rangle, \langle h, d\rangle$. Остались обратные рёбра $\langle i, d\rangle, \langle i, e\rangle$ и ребро дерева $\langle g, i\rangle$. Формируем обратные пути следующим образом. Для каждого пути выбираем рёбра дерева, соединяющие вершины, которые не принадлежат опорному циклу, и одно обратное ребро так, чтобы концевые вершины обратного маршрута принадлежали опорному циклу. В нашем случае это будет обратный путь $\langle g, e\rangle = \langle g, i\rangle + \langle i, e\rangle$ (либо можно сформировать другой обратный путь $\langle g, d\rangle = \langle g, i\rangle + \langle i, d\rangle$). Располагаем все выбранные маршруты, кроме обратного ребра $\langle i, d\rangle$, не имеющего смежной вершины с вершинами опорного цикла, внутри опорного цикла $\langle a, b, c, d, e, f, g, h, a\rangle$ (рис. 7). Формируем множество $X_2 \subseteq X$ вершин графа, не принадлежащих опорному циклу. Вершина i не принадлежит опорному циклу, поэтому включается в множество X_2.

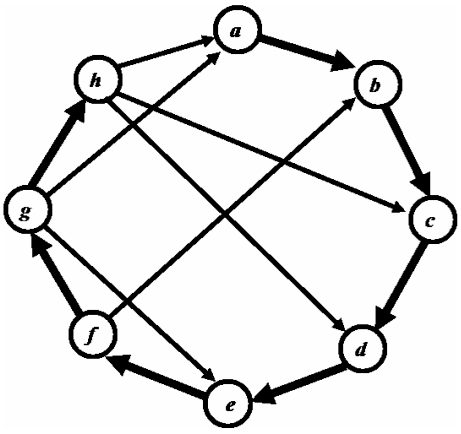

Рис. 7. Расположение обратных путей внутри опорного цикла

Определение 8 [16]. *Координатно-базисная система* — это замкнутая последовательность ориентированных дуг графа.

Для нашей задачи координатно-базисная система состоит из ориентированных дуг, принадлежащих опорному циклу. Обозначим проекцию обратного маршрута $\langle x, y\rangle$ на координатно-базисную систему как $\mathrm{pr}\langle x, y\rangle$. В нашем случае проекция маршрутов на

координатно-базисную систему имеет вид (см. рис. 7)

$$\begin{gathered}
\mathrm{pr}\langle g,a\rangle = \langle g,h\rangle + \langle h,a\rangle = \langle g,h,a\rangle;\\
\mathrm{pr}\langle f,b\rangle = \langle f,g\rangle + \langle g,h\rangle + \langle h,a\rangle + \langle a,b\rangle = \langle f,g,h,a\rangle;\\
\mathrm{pr}\langle h,c\rangle = \langle h,a\rangle + \langle a,b\rangle + \langle b,c\rangle = \langle h,a,b,c\rangle;\\
\mathrm{pr}\langle h,e\rangle = \langle h,a\rangle + \langle a,b\rangle + \langle b,c\rangle + \langle c,e\rangle = \langle h,a,b,c,e\rangle;\\
\mathrm{pr}\langle g,e\rangle = \langle g,h\rangle + \langle h,a\rangle + \langle a,b\rangle + \langle b,c\rangle + \langle c,d\rangle + \langle d,e\rangle = \langle g,h,a,b,c,d,e\rangle.
\end{gathered}$$

4. Определение пересечения обратных маршрутов

Будем рассматривать попарное пересечений маршрутов как множественное пересечение их проекций. Если пересечение двух подмножеств пусто или одно подмножество включается в другое, то обратные маршруты не пересекаются; иначе пара обратных маршрутов пересекается. Для нашего примера:

$$\begin{aligned}
\mathrm{pr}\langle g,a\rangle \cap \mathrm{pr}\langle f,b\rangle &= \langle g,h,a\rangle \subseteq \langle f,g,h,a,b\rangle - \text{пересечения нет};\\
\mathrm{pr}\langle g,a\rangle \cap \mathrm{pr}\langle h,c\rangle &= \langle g,h,a\rangle \cap \langle h,a,b,c\rangle = \{h,a\} - \text{пересечение есть};\\
\mathrm{pr}\langle g,a\rangle \cap \mathrm{pr}\langle h,d\rangle &= \langle g,h,a\rangle \cap \langle h,a,b,c,d\rangle = \{h,a\} - \text{пересечение есть};\\
\mathrm{pr}\langle g,a\rangle \cap \mathrm{pr}\langle g,e\rangle &= \langle g,h,a\rangle \subseteq \langle g,h,a,b,c,d,e\rangle - \text{пересечения нет};\\
\mathrm{pr}\langle f,b\rangle \cap \mathrm{pr}\langle h,c\rangle &= \langle f,g,h,a,b\rangle \cap \langle h,a,b,c\rangle = \{h,a,b\} - \text{пересечение есть};\\
\mathrm{pr}\langle f,b\rangle \cap \mathrm{pr}\langle h,d\rangle &= \langle f,g,h,a,b\rangle \cap \langle h,a,b,c,d\rangle = \{h,a,b\} - \text{пересечение есть};\\
\mathrm{pr}\langle f,b\rangle \cap \mathrm{pr}\langle g,e\rangle &= \langle f,g,h,a,b\rangle \cap \langle g,h,a,b,c,d,e\rangle = \{h,a,b\} - \text{пересечение есть};\\
\mathrm{pr}\langle h,c\rangle \cap \mathrm{pr}\langle h,d\rangle &= \langle h,a,b,c\rangle \subseteq \langle h,a,b,c,d\rangle - \text{пересечения нет};\\
\mathrm{pr}\langle h,c\rangle \cap \mathrm{pr}\langle g,e\rangle &= \langle h,a,b,c\rangle \subseteq \langle g,h,a,b,c,d,e\rangle - \text{пересечения нет};\\
\mathrm{pr}\langle h,d\rangle \cap \mathrm{pr}\langle g,e\rangle &= \langle h,a,b,c,d\rangle \subseteq \langle g,h,a,b,c,d,e\rangle - \text{пересечения нет}.
\end{aligned}$$

Составим таблицу пересечения маршрутов:

Путь	Пересекаются	Не пересекаются
$\langle g,a\rangle$	$\langle h,c\rangle, \langle h,d\rangle$	$\langle f,b\rangle, \langle g,e\rangle$
$\langle f,b\rangle$	$\langle h,c\rangle, \langle h,d\rangle, \langle g,e\rangle$	$\langle g,a\rangle$
$\langle h,c\rangle$	$\langle g,a\rangle, \langle f,b\rangle$	$\langle h,d\rangle, \langle g,e\rangle$
$\langle h,d\rangle$	$\langle g,a\rangle, \langle f,b\rangle$	$\langle h,d\rangle, \langle g,e\rangle$
$\langle g,e\rangle$	$\langle f,b\rangle$	$\langle g,a\rangle, \langle h,c\rangle, \langle h,d\rangle$

Используя методы алгебры логики, образуем два множества обратных маршрутов, не пересекающихся между собой: $M_1 = \{\langle g,a\rangle, \langle f,b\rangle\}$, $M_2 = \{\langle h,c\rangle, \langle h,d\rangle, \langle g,e\rangle\}$. Расположим множества маршрутов M_2 внутри опорного цикла, а множество обратных маршрутов M_1 — вне опорного цикла (рис. 8). Таким образом, поверхность разделилась на две части: внутреннюю и внешнюю.

Определение 9. *Обруч* — это простой цикл, образованный кольцевой суммой изометрических циклов, каждый из которых имеет по крайней мере одно общее ребро с другим изометрическим циклом.

Заметим, что изометрический цикл можно рассматривать как частный случай обруча. В свою очередь, обратные рёбра и обратные пути разбивают поверхности на грани, где границами граней служат обручи [12].

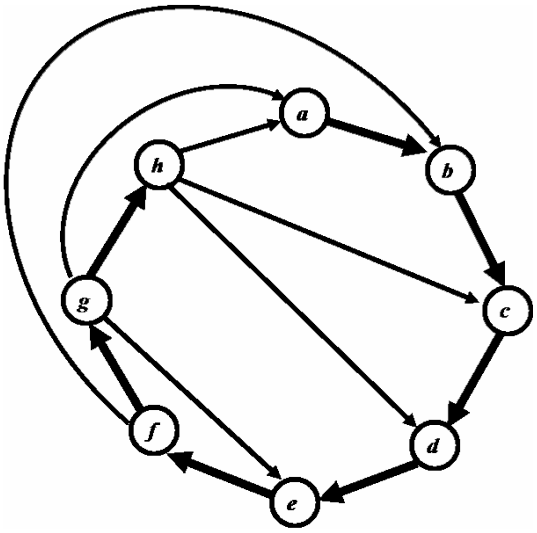

Рис. 8. Размещение обратных маршрутов внутри и вне опорного цикла

5. Нахождение обручей

Для определения обручей поступим следующим образом. Так как исходный граф неориентированный, для внутренних граней обручи будут характеризоваться расположением ориентированных рёбер по часовой стрелке (рис. 9), а для внешних граней — против часовой стрелки (рис. 10). Обратное ребро заменим неориентированным ребром или двумя разнонаправленными ориентированными дугами. Будем искать обручи.

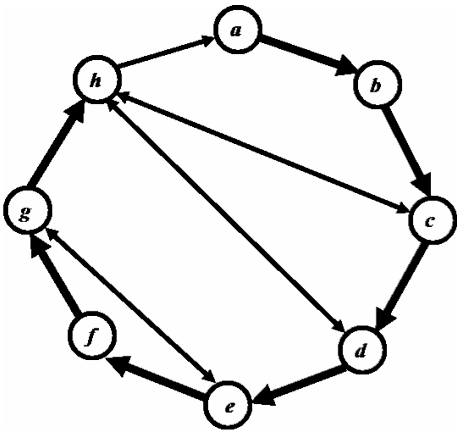

Рис. 9. Размещение обратных маршрутов внутри опорного цикла

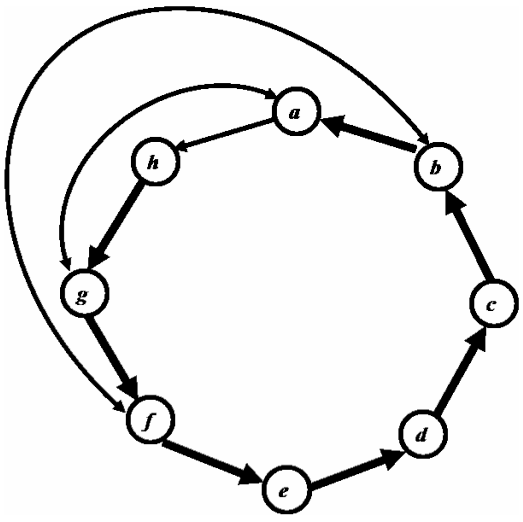

Рис. 10. Размещение обратных маршрутов вне опорного цикла

Для графа, представленного на рис. 9, запишем матрицу смежностей, сохраняя порядок следования вершин в опорном цикле:

	a	b	c	d	e	f	g	h
a		**1**						
b			**1**					
c				**1**				1
d					**1**			1
e						**1**	1	
f							**1**	
g					1			**1**
h	**1**		1	1				

Диагональ матрицы, состоящей из помеченных элементов, назовём *верхней*. Она состоит из ориентированных рёбер, принадлежащих опорному циклу.

В матрице выбираем первый встречный элемент, принадлежащий обратному ребру. Рассматривая соответствующую строку матрицы, отмечаем вершину, принадлежащую данному элементу. Переходим на верхнюю диагональ и, проходя по диагонали, последовательно находим строку, содержащую в столбце отмеченную вершину. Например, в строке с вершиной c выбираем столбец h, переходим на верхнюю диагональ и, последовательно проходя по диагонали, в строке вершины b находим столбец, содержащий вершину c. Тем самым мы сформировали обруч $\langle c, h, a, b, c\rangle$. После выделения цикла $\langle \boldsymbol{c}, \boldsymbol{h}, a, b, c\rangle$ удаляем из матрицы отмеченные элементы. В новой матрице выбираем противоположное ребро $\langle h, c\rangle$ и строим обруч $\langle h, c, d, h\rangle$:

	a	b	c	d	e	f	g	h
a								
b								
c				**1**				
d					1			**1**
e						1	1	
f							1	
g					1			1
h			**1**	1				

После выделения обруча $\langle h, c, \boldsymbol{d}, \boldsymbol{h}\rangle$ с участием обратного ребра $\langle d, h\rangle$ матрица приобретает вид, в котором отсутствуют отмеченные элементы. Выбираем противоположное ребро $\langle h, d\rangle$ и строим обруч $\langle h, d, e, f, g, h\rangle$:

	a	b	c	d	e	f	g	h
a								
b								
c								
d					**1**			
e						**1**	1	
f							**1**	
g					1			**1**
h			**1**					

После выделения граничного обруча $\langle h, d, e, f, g, h\rangle$ в матрице смежностей остается только одно неориентированное ребро ($\langle e, g\rangle$, $\langle g, e\rangle$). Находим обруч, содержащий вершины e и g — $\langle h, d, e, f, g, h\rangle$. Представляем этот обруч в виде суммы ориентированных рёбер и добавляем ориентированные рёбра ($\langle e, g\rangle + \langle g, e\rangle$):

$$\langle h, d\rangle + \langle d, e\rangle + \langle e, f\rangle + \langle f, g\rangle + \langle g, h\rangle + (\langle \boldsymbol{e}, \boldsymbol{g}\rangle + \langle \boldsymbol{g}, \boldsymbol{e}\rangle).$$

Переориентируем последовательность ориентированных рёбер и получаем два новых обруча: $\langle h, d\rangle + \langle d, e\rangle + \langle \boldsymbol{e}, \boldsymbol{g}\rangle + \langle g, h\rangle$ и $\langle e, f\rangle + \langle f, g\rangle + \langle \boldsymbol{g}, \boldsymbol{e}\rangle$.

Для графа, представленного на рис. 10, строим матрицу смежностей с верхней диагональю:

	a	h	g	f	e	d	c	b
a		**1**	1					
h			**1**					
g	1			**1**				
f					**1**			1
e						**1**		
d							**1**	
c								**1**
b	**1**			1				

В матрице выбираем первый встречный элемент, принадлежащий обратному ребру. Рассматривая соответствующую строку матрицы, отмечаем вершину, принадлежащую данному элементу. Переходим на верхнюю диагональ и, проходя по диагонали, последовательно находим строку, содержащую отмеченную вершину. Например, в строке, содержащей вершину a, выбираем столбец g, переходим на верхнюю диагональ, последовательно проходя по верхней диагонали, в строке h находим столбец, содержащий вершину a. Формируем обруч $\langle \boldsymbol{a}, \boldsymbol{g}, f, e, d, c, b, a\rangle$. Строим новую матрицу. Выбираем противоположное ребро $\langle g, a\rangle$ и строим обруч $\langle g, a, h, g\rangle$:

	a	h	g	f	e	d	c	b
a		**1**						
h			**1**					
g	**1**							
f								1
e								
d								
c								
b				1				

После выделения обруча $\langle g, a, h, g\rangle$ в матрице смежностей остаётся только одно неориентированное ребро ($\langle f, b\rangle$, $\langle b, f\rangle$). Находим обруч, содержащий вершины f и b. Это цикл $\langle a, g, f, e, d, c, b, a\rangle$. Представляем его в виде суммы ориентированных рёбер и добавляем ориентированные дуги ($\langle f, b\rangle + \langle b, f\rangle$):

$$\langle a, g\rangle + \langle g, f\rangle + \langle f, e\rangle + \langle e, d\rangle + \langle d, c\rangle + \langle c, b\rangle + \langle b, a\rangle + (\langle \boldsymbol{f}, \boldsymbol{b}\rangle + \langle \boldsymbol{b}, \boldsymbol{f}\rangle).$$

Переориентируем последовательность ориентированных рёбер и получаем два новых обруча: $\langle a, g\rangle + \langle g, f\rangle + \langle \boldsymbol{f}, \boldsymbol{b}\rangle + \langle b, a\rangle$ и $\langle f, e\rangle + \langle e, d\rangle + \langle d, c\rangle + \langle c, b\rangle + \langle \boldsymbol{b}, \boldsymbol{f}\rangle$.

Таким образом, получена система обручей, описывающая процесс разбиения пространства $\mathbb{R}^2$ на грани, причём сумма всех обручей, согласно теореме Маклейна, есть пустое множество:

$$\sum_{j=1}^{m-n+2} c_j = \varnothing.$$

Система обручей имеет следующий вид:

$$\begin{gathered}
c_1 = \langle c, h\rangle + \langle h, a\rangle + \langle a, b\rangle + \langle b, c\rangle = \langle c, h, a, b, c\rangle;\\
c_2 = \langle h, c\rangle + \langle c, d\rangle + \langle d, h\rangle = \langle h, c, d, h\rangle;\\
c_3 = \langle h, d\rangle + \langle d, e\rangle + \langle e, g\rangle + \langle g, h\rangle = \langle h, d, e, g, h\rangle;\\
c_4 = \langle e, f\rangle + \langle f, g\rangle + \langle g, e\rangle = \langle e, f, g, e\rangle;\\
c_5 = \langle g, a\rangle + \langle a, h\rangle + \langle h, g\rangle = \langle g, a, h, g\rangle;\\
c_6 = \langle a, g\rangle + \langle g, f\rangle + \langle f, b\rangle + \langle b, a\rangle = \langle a, g, f, b, a\rangle;\\
c_7 = \langle f, e\rangle + \langle e, d\rangle + \langle d, c\rangle + \langle c, b\rangle + \langle b, f\rangle = \langle f, e, d, c, b, f\rangle.
\end{gathered}$$

6. Ввод вершин, не принадлежащих опорному циклу

Введём вершину i, ранее не участвовавшую в процессе построения топологического рисунка графа. Вершина i расположена между вершинами g и e в маршруте $\langle g, e\rangle$ и в маршруте $\langle e, g\rangle$. Расположим вершину i в новых обручах c_3 и c_4:

$$c_3 = \langle h, d\rangle + \langle d, e\rangle + \langle e, i\rangle + \langle i, g\rangle + \langle g, h\rangle = \langle h, d, e, i, g, h\rangle;$$
$$c_4 = \langle e, f\rangle + \langle f, g\rangle + \langle g, i\rangle + \langle i, e\rangle = \langle e, f, g, i, e\rangle.$$

Введём ребро ($\langle i, d\rangle + \langle d, i\rangle$), ранее не участвовавшее в процессе построения топологического рисунка, в обруч c_3, охватывающий вершины i и d:

$$c_3 = \langle h, d\rangle + \langle d, e\rangle + \langle e, i\rangle + \langle i, g\rangle + \langle g, h\rangle + (\langle \boldsymbol{i}, \boldsymbol{d}\rangle + \langle \boldsymbol{d}, \boldsymbol{i}\rangle).$$

Переориентируем последовательность ориентированных рёбер и получим два новых обруча:

$$c_3 = \langle h, d\rangle + \langle \boldsymbol{d}, \boldsymbol{i}\rangle + \langle i, g\rangle + \langle g, h\rangle = \langle h, d, i, g, h\rangle;$$
$$c_8 = \langle e, i\rangle + \langle \boldsymbol{i}, \boldsymbol{d}\rangle + \langle d, e\rangle = \langle e, i, d, e\rangle.$$

Все обратные пути и обратные рёбра построены. Добавляем обод, рассчитанный по формуле $c_0 = -\sum\limits_{i=1}^{m-n+1} c_i$. Полученная система изометрических циклов формирует диаграмму вращения вершин и описывает топологический рисунок плоского графа (рис. 11).

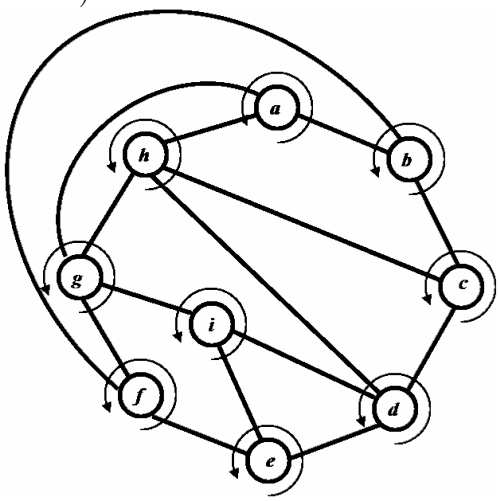

Диаграмма вращения вершин

$\hbar(a):$	h	b	g	
$\hbar(b):$	a	c	f	
$\hbar(c):$	h	d	b	
$\hbar(d):$	h	i	e	c
$\hbar(e):$	f	d	i	
$\hbar(f):$	g	b	e	
$\hbar(g):$	a	f	i	h
$\hbar(h):$	g	d	c	a
$\hbar(i):$	g	e	d	

Рис. 11. Топологический рисунок графа

7. Алгоритм проверки планарности и построения топологического рисунка планарного графа (поиском в глубину)

Опишем алгоритм проверки планарности и построения топологического рисунка планарного графа. Вычислительную сложность алгоритма представим как сумму элементарных процессов в зависимости от количества рёбер графа. Сложность построения подмножества хорд графа можно определить функцией $f_1(m) = m-n+1$; процесса парного пересечения хорд — функцией $f_2(m) = (m-n+1)(m-n)/2$; вычислительную сложность выделения изометрических циклов можно определить как удвоенное количество хорд и описать функцией $f_3(m) = 2(m-n+1)$. В результате получим вычислительную сложность алгоритма $\mathrm{O}(m^2)$.

Алгоритм 1. Проверка планарности и построение топологического рисунка планарного графа

1: *Инициализация.* Задан пронумерованный несепарабельный граф G.
2: *Построение DFS-дерева.* Поиском в глубину выделяем DFS-дерево графа.
3: *Выделение опорного цикла.* Строим матрицу фундаментальных циклов графа и выделяем цикл максимальной длины — опорный цикл.
4: *Формирование обратных путей.* Разбиваем множество вершин на подмножества X_1 и X_2 — включённых и не включённых в опорный цикл. Формируем обратные пути из матрицы фундаментальных циклов путём соединения хорды и элементов соответствующей строки, состоящей только из вершин подмножества X_2.
5: *Обратный путь — петля?* Если имеется обратный путь, состоящий только из вершин подмножества X_1, то этот обратный путь — петля, идём на шаг 3.
6: *Определение проекций обратных путей.* Проекции обратных путей состоят из элементов соответствующей строки хорды в матрице фундаментальных циклов, принадлежащих вершинам подмножества X_1.
7: *Определение минимальной длины проекции обратного пути.* Если нет проекции обратного пути длины, меньшей самого обратного пути, то идём на шаг 9.
8: *Перестройка DFS-дерева.* Включаем вершины выбранного обратного пути в опорный цикл и перестраиваем DFS-дерево. Если необходимо — удаляем вершины из опорного цикла. Параллельно перестраивается матрица фундаментальных циклов и подмножества вершин X_1 и X_2; идём на шаг 5.
9: *Определение пересечения обратных путей.* Распределяем обратные пути по группам. Определяем проекцию обратных путей и обратных рёбер на опорный цикл. Попарно просматривая множественное пересечение проекций обратных маршрутов, определяем их пересечение.
10: *Формирование непересекающихся подмножеств обратных маршрутов M_1 и M_2.* Если сформировать такие множества нельзя, граф не планарен, идём на шаг 15.
11: *Выделение обручей.* Методом, основанным на выделении ориентированных циклов в упорядоченной матрице смежностей, формируем систему обручей.
12: *Введение вершин из подмножества X_2.* Располагаем в соответствующих обручах в определённом порядке вершины из подмножества X_2, ранее не участвовавшие в процессе построения топологического рисунка графа.
13: *Проведение оставшихся обратных рёбер.* Строим изометрические циклы, помещая вновь введённые обратные рёбра внутрь соответствующих обручей с проверкой на пересечение. Если имеется пересечение, то граф непланарный, идём на шаг 15.
14: *Построение рисунка планарного графа.* Для выделенного базиса подпространства циклов, состоящего из изометрических циклов и обода, формируем диаграмму вращений вершин графа и получаем топологический рисунок плоского графа.
15: *Конец работы алгоритма.*

Заключение

В работе представлен алгоритм проверки планарности графа с одновременным построением математических структур для описания топологического рисунка графа с целью его визуализации. Показано, что математическими структурами для построения топологического рисунка графа являются изометрические циклы и вращение вершин графа и что вращение вершин плоского графа индуцирует базис подпространства циклов, состоящий из изометрических циклов.

Отличительной особенностью представленного метода является введение операции определения пересечения рёбер как пересечения их проекций на координатно-базисную систему, в качестве которой выступает опорный цикл DFS-дерева графа. Введение опорного цикла разбивает оставшуюся часть графа на множество обратных маршрутов, а применение операции проверки рёбер на пересечение позволяет распределить обратные маршруты на те, которые находятся внутри опорного цикла, и вне его. С целью получения минимального количества обратных путей производится перестройка DFS-дерева графа для выделения опорного цикла большей длины. Обратные пути совместно с рёбрами опорного цикла образуют обручи графа. В свою очередь, обручи позволяют выделить изометрические циклы графа. Выделенные изометрические циклы, удовлетворяющие условию теоремы Маклейна, индуцируют вращение вершин графа, тем самым порождая топологический рисунок плоской части графа. Топологический рисунок позволяет описывать процесс планаризации алгебраическими методами, не производя никаких геометрических построений на плоскости. Получение вращения вершин графа сразу решает две важнейшие задачи теории графов: проверки на планарность и построения топологического рисунка плоского графа. Построение топологического рисунка плоского графа является основой для дальнейшего построения топологического рисунка непланарного графа с заданными характеристиками, а также служит основой для геометрического построения рисунка графа (визуализации).

К недостаткам алгоритма можно отнести его сложность. Сегодня существуют эффективные алгоритмы, позволяющие определить, является ли граф планарным, с линейной от количества вершин графа сложностью. Предложенный в работе алгоритм обладает сложностью $O(m^2)$. В то же время, в отличие от алгоритмов с линейной сложностью, предложенный алгоритм позволяет получить топологический рисунок графа, который может затем использоваться для визуализации графа. В свою очередь, визуализация планарных графов является важнейшей подзадачей при проектировании сложных изделий и систем, плоских конструктивов, анализе социальных сетей, а также при решении других актуальных прикладных задач. Важно отметить, что использование топологического подхода позволяет создавать многовариантные системы автоматизированного проектирования плоских конструктивов и других сложных систем, так как многовариантность построения множества плоских частей непланарного графа позволяет получить множество решений. Так, на основе понятия топологического рисунка графа и предложенного в данной работе алгоритма авторами разработан топологический подход к проведению соединений в плоских конструктивах [2]. В дальнейшем планируется развивать предложенный подход для полного решения задачи построения общего рисунка всех соединений в плоском конструктиве и нахождения оптимального решения с учётом конструкторских и технологических требований.

Визуализация планарных графов играет важную роль в различных сферах научных исследований и актуальна для многих прикладных областей. Глубокие теоретические результаты, полученные в дискретной математике и в теории графов, привели к созданию быстрых алгоритмов для решения некоторых частных задач, связанных с визуализацией планарных графов. Однако развитие различных прикладных задач, таких, например, как анализ больших сетей и больших объёмов данных, а также проектирование сложных инженерных систем (плоские конструктивы в радиоэлектронике, комплексы транспортно-инженерных коммуникаций крупных предприятий, логистические системы) требуют новых эффективных подходов к их решению. По мнению авторов, среди наиболее важных направлений исследований, связанных с визуализа-

цией графов, следует отметить выделение плоской части в непланарных графах и размещение непланарных графов на плоскости, многовариантность и оптимизацию размещений по заданным критериям, создание эффективных алгоритмов визуализации при наличии ограничений. Состояние исследований и нерешённые задачи в области визуализации планарных графов проанализированы подробнее в работе [6].

ЛИТЕРАТУРА

1. *Курапов С. В., Давидовский М. В.* Два подхода к проведению соединений в плоских конструктивах // Компоненты и технологии. 2015. № 7. С. 142–147.
2. *Курапов С. В., Давидовский М. В.* Топологический подход к проведению соединений в плоских конструктивах // Компоненты и технологии. 2015. № 11. С. 127–130.
3. *Апанович З. В.* От рисования графов к визуализации информации // Препринт № 148. Новосибирск: ИСИ СО РАН, 2007. 16 с.
4. *Di Battista G., Eades P., Tamassia R., and Tollis I. G.* Algorithms for drawing graphs: an annotated bibliography // Comp. Geom. 1994. V. 4. No. 5. P. 235–282.
5. *Tamassia R.* Handbook of Graph Drawing and Visualization. Boca Raton: Chapman and Hall/CRC, 2013. 844 p.
6. *Курапов С. В., Толок А. В.* Методы построения топологического рисунка графа // Автоматика и телемеханика. 2013. № 9. С. 78–97.
7. *Kavitha T., Liebchen C., Mehlhorn K., et al.* Cycle bases in graphs — characterization, algorithms, complexity, and applications // Comput. Sci. Rev. 2009. No. 3. P. 199–243.
8. *Деза М., Гришухин В. П., Штогрин М. И.* Изометрические полиэдральные подграфы в гиперкубах и кубических решетках. М.: МЦНМО, 2007. 192 с.
9. *Рингель Г.* Теорема о раскраске карт. М.: Мир, 1977. 126 с.
10. *Зыков А. А.* Теория конечных графов. Новосибирск: Наука, 1969. 554 с.
11. *Свами М., Тхуласираман К.* Графы, сети и алгоритмы. М.: Мир, 1984. 455 с.
12. *Харари Ф.* Теория графов. М.: Мир, 1973. 300 с.
13. *Мак-Лейн С.* Комбинаторное условие для плоских графов // Кибернетический сборник. Новая серия. 1970. № 7. С. 68–77.
14. *Хопкрофт Дж. Е., Тарьян Р. Е.* Изоморфизм планарных графов // Кибернетический сборник. Новая серия. 1975. № 12. С. 39–61.
15. *Рейнгольд Э., Нивергельт Ю., Део Н.* Комбинаторные алгоритмы. Теория и практика. М.: Мир, 1980. 480 с.
16. *Раппопорт Л. И., Мороговский Б. Н., Поливцев С. А.* Векторная алгебра пересечений // Многопроцессорные вычислительные структуры. 1982. № 2(11). С. 53–56.

REFERENCES

1. *Kurapov S. V. and Davidovskiy M. V.* Dva podkhoda k provedeniyu soedineniy v ploskikh konstruktivakh [Two approaches to connections conducting in flat form factor]. Komponenty i Tekhnologii, 2015, no. 7, pp. 142—147. (in Russian)
2. *Kurapov S. V. and Davidovskiy M. V.* Topologicheskiy podkhod k provedeniyu soedineniy v ploskikh konstruktivakh [Topological approach to connections conducting in flat form factor]. Komponenty i Tekhnologii, 2015, no. 11, pp. 127–130. (in Russian)
3. *Apanovich Z. V.* Ot risovaniya grafov k vizualizatsii informatsii [From Graphs Drawing to Information Visualization]. Preprint no. 148. Novosibirsk, IIS RAS SB Publ., 2007. 16 p. (in Russian)
4. *Di Battista G., Eades P., Tamassia R., and Tollis I. G.* Algorithms for drawing graphs: an annotated bibliography. Comp. Geom., 1994, vol. 4, no. 5, pp. 235–282.

5. *Tamassia R.* Handbook of Graph Drawing and Visualization. Boca Raton: Chapman and Hall/CRC, 2013. 844 p.
6. *Kurapov S. V. and Tolok A. V.* The topological drawing of a graph: Construction methods. Automation and Remote Control, 2013, vol. 74, iss. 9, pp. 1494–1509.
7. *Kavitha T., Liebchen C., Mehlhorn K., et al.* Cycle bases in graphs — characterization, algorithms, complexity, and applications. Comput. Sci. Rev., 2009, no. 3, pp. 199–243.
8. *Deza M., Grishukhin V. P., and Shtogrin M. I.* Izometricheskie poliedral'nye podgrafy v giperkubakh i kubicheskikh reshetkakh [Isometric Polyhedral Subgraphs in Hypercubes and Cubic Lattices]. Moscow, MCCME Publ., 2007. 192 p. (in Russian)
9. *Ringel' G.* Teorema o raskraske kart [Theorem of Maps Coloring]. Moscow, Mir Publ., 1977. 126 p. (in Russian)
10. *Zykov A. A.* Teoriya konechnykh grafov [Finite Graphs Theory]. Novosibirsk, Nauka Publ., 1969. 554 p. (in Russian)
11. *Swamy M. N. S. and Thulasiraman K.* Graphs, Networks and Algorithms. Wiley, 1980. 612 p.
12. *Harary F.* Graph Theory. Addison–Wesley, 1969.
13. *Mak-Leyn S.* Kombinatornoe uslovie dlya ploskikh grafov [A combinatorial condition for planar graphs.] Kiberneticheskiy Sbornik. Novaya Seriya, 1970, no. 7, pp. 68–77. (in Russian)
14. *Khopkroft Dzh. E. and Tar'yan R. E.* Izomorfizm planarnykh grafov [The isomorphism of planar graphs]. Kiberneticheskiy Sbornik. Novaya seriya, 1975, no. 12, pp. 39–61. (in Russian)
15. *Reingold E. M., Nievergelt J., and Deo N.* Combinatorial Algorithms: Theory and Practice. Prentice Hall College Div, 1977.
16. *Rappoport L. I., Morogovskiy B. N., and Polivtsev S. A.* Vektornaya algebra peresecheniy [Vector algebra of intersections.] Mnogoprotsessornye vychislitel'nye struktury. Taganrog, TREI Publ., 1980, iss. 2(11), pp. 53–56. (in Russian)

УДК 519.17

О КОЛИЧЕСТВЕ ШПЕРНЕРОВЫХ ВЕРШИН В ДЕРЕВЕ

В. Н. Салий

Саратовский государственный университет им. Н. Г. Чернышевского, г. Саратов, Россия

Вершина v дерева T называется шпернеровой вершиной, если входящее дерево $T(v)$, полученное из T ориентацией всех рёбер в направлении к v, обладает шпернеровым свойством: в нём среди наибольших (по числу элементов) подмножеств, состоящих из попарно недостижимых вершин, по крайней мере в одном все вершины равноудалены от v. Приводятся явные способы подсчёта количества шпернеровых вершин в деревьях некоторых типов.

Ключевые слова: *дерево, шпернерова вершина, цепь, звезда, пальма, шеренга, гусеница, кортеж пальм.*

DOI 10.17223/20710410/32/8

ON THE NUMBER OF SPERNER VERTICES IN A TREE

V. N. Salii

Saratov State University, Saratov, Russia

E-mail: SaliiVN@info.sgu.ru

A vertex v of a tree T is called a Sperner vertex if the in-tree $T(v)$ obtained from T by orientation of all edges towards v has the Sperner property, i. e. there exists a largest subset A of mutually unreachable vertices in it such that all vertices in A are equidistant to v. Some explicit methods to count the number of Sperner vertices in certain special trees are presented.

Keywords: *graph, Sperner vertex, path, star, palm-tree, rank, caterpillar, train of palm-trees.*

Пусть $G = (V, \alpha)$ — бесконтурный (ориентированный) граф с множеством вершин V и отношением смежности вершин α. *Антицепью* в G называется набор вершин $A \subseteq V$, такой, что никакая вершина, принадлежащая A, недостижима из других вершин этого множества. Например, антицепью является совокупность всех источников графа G, т. е. его вершин, недостижимых из других вершин. Антицепи с наибольшим числом вершин по определению являются главными. Под *высотой вершины* $v \in V$ понимается наибольшая из длин цепей в G, началом которых служит v. Например, все стоки графа G, т. е. вершины, из которых недостижимы другие вершины, имеют высоту 0. Антицепь A будем называть *правильной*, если она состоит из вершин с одинаковой высотой. Говорят, что граф G *обладает шпернеровым свойством*, или что он является *шпернеровым графом*, если среди его главных антицепей есть хотя бы одна правильная. Это равносильно тому, что множество V вершин графа G, упорядоченное отношением достижимости вершин, является шпернеровым упорядоченным множеством. Определяющее свойство для антицепей в конечных упорядоченных множествах впервые рассмотрел Е. Шпернер в 1928 г. [1]. С тех пор оно интенсивно изучается в различных конкретных ситуациях (см., например, ссылки в [2]).

Пусть $T = (V, \alpha)$ — некоторое дерево (неориентированный связный граф без циклов) и $v \in V$ — одна из его вершин. Любая другая вершина u дерева T связана с v единственной цепью. Ориентируя рёбра всех подобных цепей в направлении к v, получим бесконтурный граф с единственным стоком v — входящее дерево $T(v)$ с корнем v. Если $T(v)$ окажется шпернеровым графом, вершина v называется шпернеровой.

Сколько шпернеровых вершин может иметь произвольное дерево?

Вершины всякого дерева по своим степеням разбиваются на три класса: висячие (или листья) имеют степень 1, у проходных вершин степень равна 2, у точек ветвления она не менее 3. *Типом вершины* v в дереве T назовём пару $t(v) = (h(v), b(v))$, где $h(v)$ — наименьшее из расстояний от v до отличных от v висячих вершин; $b(v)$ — наибольшее из расстояний от v до отличных от v точек ветвления.

Теорема 1 (см. также [3, теорема 2]). Вершина v дерева T шпернерова тогда и только тогда, когда $h(v) > b(v)$.

Доказательство. Необходимость. Пусть v — шпернерова вершина. Тогда во входящем дереве $T(v)$ существует правильная главная антицепь A, все вершины которой имеют одну и ту же высоту $k > 0$. Если $h(v) < k$, то ближайшая к v висячая вершина u дерева T не входит в A. Но если это так, то, присоединив к главной антицепи A вершину u, получим в $T(v)$ антицепь большей, чем у A, длины, что невозможно. Значит, $h(v) \geqslant k$. Допустим, что $b(v) \geqslant h(v)$. Это означает, что в T есть точка ветвления w, удалённая от v не менее чем на k, т. е. имеющая в $T(v)$ высоту $\geqslant k$. Вершина w достижима в T не менее чем из двух висячих вершин. При этом w либо сама входит в антицепь A, либо в A имеется единственная вершина, достижимая из w. Заменив эту вершину (или саму w, если $w \in A$) парой висячих вершин, из которых достижима w, получим антицепь большей, чем у A, длины, что невозможно. Значит, $b(v) < k$ и $h(v) > b(v)$.

Достаточность. Пусть для вершины v дерева T выполняется неравенство $h(v) > b(v)$. Обозначим через A совокупность всех вершин дерева $T(v)$, имеющих высоту $k = h(v)$. Так как все точки ветвления лежат в $T(v)$ ниже уровня k, каждая вершина, входящая в антицепь A, достижима в точности из одной висячей вершины дерева T. Значит, количество вершин в A равно количеству висячих вершин дерева T и, следовательно, A имеет наибольшее возможное для антицепей в $T(v)$ количество элементов, т. е. A — главная в $T(v)$ антицепь. При этом она составлена из вершин с одинаковой высотой, т. е. является правильной. ■

В [3] предложен полиномиальный алгоритм для проверки свойства шпернеровости у произвольной предъявленной вершины дерева. Разумеется, его можно использовать и для установления количества $s(T)$ шпернеровых вершин в дереве T. Вместе с тем представляют интерес и явные способы вычисления величины $s(T)$ для деревьев того или иного частного вида. Приведём некоторые примеры таких расчётов.

1. Цепь (path) $P_n = v_0 v_1 \dots v_n$. Здесь нет точек ветвления и каждая вершина является шпернеровой: $s(P_n) = n + 1$.

2. Звезда (star) S_n, $n \geqslant 3$, — дерево с единственной точкой ветвления (центр звезды) v_0 и n висячими вершинами (лучи) $v_1, v_2, \dots, v_n$. Определив типы лучевых вершин $t(v_i) = (2, 1)$, $1 \leqslant i \leqslant n$, убеждаемся, что все они являются шпернеровыми. Входящее дерево $T(v_0)$ имеет правильную главную антицепь, состоящую из лучевых вершин, так что и v_0 — шпернерова, и, следовательно, $s(S_n) = n + 1$.

3. Пальма (palm-tree) $PT_{(l,n)}$, $l \geqslant 2$, $n \geqslant 2$ — цепь $u_0 u_1 \dots u_l$ (ствол), концевая вершина которой u_l (верхушка) является центром звезды с лучами (листьями) $\lambda_1, \lambda_2, \dots, \lambda_n$.

Для корня u_0 имеем $t(u_0) = (l+1, l)$, для листьев $t(\lambda_i) = (2, 1)$, $1 \leqslant i \leqslant n$, для верхушки $h(u_l) = 1$, так что все эти $n+2$ вершин — шпернеровы. Для проходных вершин u_i, $0 < i < l$, ствола имеем $b(u_i) = l - i$, $h(u_i) = \min(i, l+1-i)$. При $i \geqslant l+1-i$, т. е. $i \geqslant [(l+1)/2]$, вершина u_i будет шпернеровой. Таких вершин в стволе имеется $[(l-1)/2]$, так что $s(PT_{(l,n)}) = n + 2 + [(l-1)/2]$.

4. Шеренга (rank) $R = (P^0(v_0^0), P^1(v_0^1), \ldots, P^n(v_0^n); P_n = v_0^0 v_0^1 \ldots v_0^n)$ — объединение некоторого количества $n \geqslant 2$ цепей с выделенными (в каждой по одной) концевыми вершинами, в свою очередь образующими цепь.

Теорема 2. Для любого целого $k \geqslant 1$ существует шеренга, имеющая в точности k шпернеровых вершин.

Доказательство. Построим шеренгу R из шести цепей $P^0 = v_0^0 v_1^0$, $P^1 = v_0^1 v_1^1 v_2^1$, $P^2 = v_0^2 v_1^2 \ldots v_l^2$, $P^3 = v_0^3 v_1^3 v_2^3$, $P^4 = v_0^4 v_1^4$, $P^5 = v_0^5 v_1^5$. Для заданного $k \geqslant 1$ положим $l = 2k - 1$. Подсчитаем типы вершин. Для листьев, отличных от v_l^2, получаем: $t(v_1^0) = (4, 5)$, $t(v_2^1) = (4, 5)$, $t(v_2^3) = (4, 4)$, $t(v_1^4) = (3, 4)$, $t(v_1^5) = (3, 5)$ — здесь нет шпернеровых вершин. Для вершин из базисной цепи, отличных от v_0^2, имеем: $t(v_0^0) = (1, 4)$, $t(v_0^1) = (2, 3)$, $t(v_0^3) = (2, 2)$, $t(v_0^4) = (1, 3)$, $t(v_0^5) = (1, 4)$ — и здесь нет шпернеровых вершин. В цепи P^2 при $k = 1$ шпернеровой будет только вершина v_1^2 с $t = (4, 3)$, так как $t(v_0^2) = (1, 2)$. Если $k \geqslant 2$, то шпернеровыми в P^2 будут k вершин $v_{2k-1}^2, v_{k-2}^2, \ldots, v_0^2$, поскольку $t(v_{2k-1}^2) = (2k+2, 2k+1)$, $t(v_{k-2}^2) = (k+1, k)$, $\ldots$, $t(v_0^2) = (3, 2)$. Если же $i \geqslant k-1$, то $h(v_i^2) = 2k-1-i \leqslant i+2 = b(v_i^2)$, и значит, такие вершины — не шпернеровы. В итоге $s(R) = k$. ■

5. Гусеница (caterpillar) — объединение C некоторого количества $k \geqslant 2$ звезд, центры которых образуют (базисную) цепь. Так как у каждого луча u будет $h(u) = 2 \leqslant b(u)$ и у центра v каждой звезды $h(v) = 1 < b(v)$, то шпернеровых вершин гусеница не имеет: $s(C) = 0$.

6. Кортеж пальм (train of palms) — объединение TP некоторого количества $n \geqslant 2$ пальм, корни которых образуют (базисную) цепь.

Теорема 3. Для любого целого $k \geqslant 0$ существует кортеж пальм, имеющий в точности k шпернеровых вершин.

Доказательство. Составим кортеж TP_l из шести пальм: $PT^0_{(1,2)}$, $PT^1_{(3,2)}$, $PT^2_{(l,2)}$, $PT^3_{(3,2)}$, $PT^4_{(1,2)}$, $PT^5_{(1,2)}$ с корнями v_0^i, $0 \leqslant i \leqslant 5$, образующими базисную цепь P_5. При вычислении величины $b(v)$ в качестве наиболее удалённой от вершины v точки ветвления выступает v_1^0 или v_1^5. Для всех листьев $h = 2 < b$, и они не шпернеровы. Для корневых вершин, отличных от v_0^2, имеем: $t(v_0^0) = (2, 6)$, $t(v_0^1) = (4, 5)$, $t(v_0^3) = (4, 4)$, $t(v_0^4) = (2, 5)$, $t(v_0^5) = (2, 6)$ — и эти вершины не шпернеровы, также как и верхушки пальм, для которых $h = 1$. При $l = 3$ получаем: $t(v_0^2) = (4, 4)$, $t(v_1^2) = (3, 5)$, $t(v_2^2) = (2, 6)$, и значит, $s(TP_3) = 0$. Если же $l = 2k + 3$, $k \geqslant 1$, то в пальме PT^2 шпернеровыми будут вершины v_j^2, $0 \leqslant j \leqslant k-1$, так как здесь $b = j + 4$ и $h = \min(j+5, 2k+4-j) = j+5$, и не будут таковыми вершины v_j^2 при $k \leqslant j \leqslant l-1$: для них $h = \min(j+5, 2k+4-j) = 2k+4-j \leqslant j+4 = b$. В итоге при $k \geqslant 0$ получается $s(TP_{2k+3}) = k$. ■

ЛИТЕРАТУРА

1. *Sperner E.* Ein Satz uber Untermengen einer endlichen Menge // Math. Zeitschrift. 1928. V. 27. Nu. 1. S. 544–548.
2. *Салий В. Н.* Шпернерово свойство для многоугольных графов // Прикладная дискретная математика. Приложение. 2014. № 7. С. 135–137.

3. *Салий В. Н.* Шпернеровы деревья // Прикладная дискретная математика. Приложение. 2015. № 8. С. 124–127.

REFERENCES

1. *Sperner E.* Ein Satz uber Untermengen einer endlichen Menge. Math. Zeitschrift, 1928, vol. 27, nu. 1, s. 544–548. (in German)
2. *Saliy V. N.* Shpernerovo svoystvo dlya mnogougol'nykh grafov [The Sperner property for polygonal graphs]. Prikladnaya diskretnaya matematika. Prilozhenie, 2014, no. 7, pp. 135–137. (in Russian)
3. *Saliy V. N.* Shpernerovy derev'ya [The Sperner property for trees]. Prikladnaya diskretnaya matematika. Prilozhenie, 2015, no. 8, pp. 124–127. (in Russian)

МАТЕМАТИЧЕСКИЕ ОСНОВЫ ИНФОРМАТИКИ И ПРОГРАММИРОВАНИЯ

УДК 510.52

О ГЕНЕРИЧЕСКОЙ СЛОЖНОСТИ ПРОБЛЕМЫ ОБЩЕЗНАЧИМОСТИ БУЛЕВЫХ ФОРМУЛ[1]

А. Н. Рыбалов

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

Генерический подход к алгоритмическим проблемам предложен А. Мясниковым, И. Каповичем, П. Шуппом и В. Шпильрайном в 2003 г. В рамках этого подхода рассматривается поведение алгоритмов на множествах почти всех входов. В данной работе изучается генерическая сложность проблемы общезначимости (тождественной истинности) булевых формул. Доказывается, что эта проблема неразрешима за полиномиальное время на любом полиномиальном строго генерическом множестве формул при условии её трудноразрешимости в худшем случае.

Ключевые слова: *генерическая сложность, проблема общезначимости булевых формул.*

DOI 10.17223/20710410/32/9

ON GENERIC COMPLEXITY OF THE VALIDITY PROBLEM FOR BOOLEAN FORMULAS

A. N. Rybalov

Sobolev Institute of Mathematics SB RAS, Novosibirsk, Russia

E-mail: alexander.rybalov@gmail.com

Generic-case approach to algorithmic problems was suggested by Miasnikov, Kapovich, Schupp and Shpilrain in 2003. This approach studies behavior of an algorithm on typical (almost all) inputs and ignores the rest of inputs. In this paper, we consider generic complexity of the validity problem for Boolean formulas and prove that this problem is generically hard if it is hard in the worst case.

Keywords: *generic complexity, validity problem for Boolean formulas.*

Введение

В работе [1] развита теория генерической сложности вычислений. 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,

[1]Работа поддержана грантом РФФИ № 16-01-00577.

быстро решающие 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 [1, 2] доказано, что таким поведением обладают многие алгоритмические проблемы алгебры, а в [3] построено генерическое множество, на котором разрешима классическая проблема остановки для машин Тьюринга с лентой, бесконечной в одном направлении. Для многих классических NP-полных проблем существуют полиномиальные генерические алгоритмы [4].

Проблема общезначимости булевых формул состоит в следующем: для любой булевой формулы, записанной в стандартном базисе $\{\vee, \wedge, \neg\}$, определить, является ли она тождественно истинной. Из классического результата С. Кука [5] о NP-полноте проблемы выполнимости следует, что проблема общезначимости является полной относительно полиномиальной сводимости в классе co-NP (он состоит из множеств, являющихся дополнениями к множествам из класса NP). Это означает, что, при условии неравенства классов P и NP, для неё не существует полиномиального алгоритма, решающего её на всём множестве булевых формул. Поэтому возникает вопрос об изучении подпроблем проблемы общезначимости и построению для них эффективных разрешающих алгоритмов. Естественным желанием является то, чтобы в эти классы попадало как можно больше формул, а в идеале «почти все» формулы. В терминах теории генерической сложности речь идёт об алгоритмах, работающих быстро на генерических множествах формул.

В данной работе доказывается, что проблема общезначимости булевых формул неразрешима за полиномиальное время на любом полиномиальном строго генерическом множестве формул при условии несовпадения классов P и NP и совпадения классов P и BPP. Здесь класс BPP — это класс проблем, разрешимых за полиномиальное время на вероятностных машинах Тьюринга. Большинство исследователей сейчас считает, что имеет место равенство $\mathrm{P} = \mathrm{BPP}$. Это равенство означает, что любой полиномиальный вероятностный алгоритм можно эффективно дерандомизировать, т. е. построить полиномиальный детерминированный алгоритм, решающий ту же задачу. Хотя это равенство пока ещё не доказано, имеются серьёзные результаты в его пользу [6]. При доказательстве основного результата работы использованы методы, развитые в [7, 8].

1. Определения

Следуя [1], дадим основные определения теории генерической сложности вычисления. Пусть I — множество всех входов, а I_n — множество входов размера n. Для любого подмножества $S \subseteq I$ определим следующую последовательность

$$\rho_n(S) = \frac{|S \cap I_n|}{|I_n|}, \ n = 1, 2, 3, \ldots$$

Величина $\rho_n(S)$ — это вероятность получить вход из множества S при случайной и равномерной генерации входов из I_n. *Асимптотической плотностью* S назовем следующий предел (если он существует):

$$\rho(S) = \lim_{n \to \infty} \rho_n(S).$$

Множество S называется *генерическим*, если $\rho(S) = 1$, и *пренебрежимым*, если $\rho(S) = 0$. Очевидно, что S генерическое тогда и только тогда, когда его дополнение $I \setminus S$ пренебрежимо.

Следуя [1], назовём множество S *строго пренебрежимым*, если последовательность $\rho_n(S)$ экспоненциально быстро сходится к нулю, т. е. существуют константы $0 < \sigma < 1$ и $C > 0$, такие, что для любого n

$$\rho_n(S) < C\sigma^n.$$

Теперь S называется *строго генерическим*, если его дополнение $I \setminus S$ строго пренебрежимо.

Множество $S \subseteq I$ *генерически разрешимо за полиномиальное время*, если существует множество $G \subseteq I$, такое, что

1) G генерическое;
2) G разрешимое за полиномиальное время;
3) $S \cap G$ разрешимое за полиномиальное время.

Если G строго генерическое, то S называется *строго генерически разрешимым за полиномиальное время*. Генерический алгоритм $\mathcal{A}$ для S работает на входе $x \in I$ следующим образом. Сначала $\mathcal{A}$ решает, принадлежит ли x множеству G. Если $x \in G$, то $\mathcal{A}$ может решить S на G, иначе $\mathcal{A}$ отвечает «Я НЕ ЗНАЮ!» Таким образом, $\mathcal{A}$ корректно решает S на «почти всех» входах (входах из генерического множества).

Имеется существенное различие между генерически разрешимыми проблемами и строго генерически разрешимыми проблемами. Допустим, имеется проблема S, разрешимая на некотором разрешимом генерическом множестве G, для которого

$$\frac{|G \cap I_n|}{|I_n|} = \frac{n-1}{n}.$$

Таким образом G — генерическое, но не строго генерическое множество. Теперь хоть проблема S и разрешима для почти всех входов, тем не менее есть быстрый способ получить «плохой» вход, на котором генерический алгоритм не работает. Быстрый (полиномиальный) алгоритм для генерации плохих входов следующий:

1) сгенерировать равномерно случайный вход x размера n;
2) если $x \in G$, повторить шаг 1, иначе закончить.

Действительно, вероятность получить только хорошие входы за n^2 раундов

$$\left(\frac{n-1}{n}\right)^{n^2} = \left(\left(1 - \frac{1}{n}\right)^n\right)^n \to e^{-n},$$

поэтому с вероятностью, очень близкой к 1, будет получен плохой вход. С другой стороны, легко видеть, что если проблема разрешима на строго генерическом множестве, то такой простой алгоритм генерации потребует экспоненциального числа раундов и будет неэффективным. Для приложений к криптографии это означает, что просто генерическая легкоразрешимость проблемы не делает эту проблему бесполезной для создания на её основе криптосистемы, так как для неё существует эффективная процедура генерации трудных входов. В то же время строго генерически легкоразрешимые проблемы в этом смысле бесполезны для криптографии.

2. Представление булевых формул

Классическое представление булевых формул с помощью таблиц истинности с практической точки зрения является громоздким в том смысле, что размер таблицы истинности растет экспоненциально с ростом числа переменных. Гораздо более компактным и практичным является представление формул с помощью бинарных деревьев. Оно часто используется в программировании различных приложений, связанных с символьными вычислениями. Кроме того, оно удобно для различного рода подсчётов.

Пусть ϕ — булева формула в базисе $\{\vee, \wedge, \neg\}$. Без ограничения общности можно считать, что в ней отрицания находятся только над переменными. Любую булеву формулу можно легко привести к такому виду с помощью законов де Моргана, поэтому в дальнейшем будем рассматривать только такие формулы. Естественным образом формуле ϕ можно сопоставить бинарное дерево T_ϕ, которое представляет конструкцию ϕ из переменных и их отрицаний с помощью конъюнкций и дизъюнкций. Внутренние вершины T_ϕ помечены символами $\vee$ и $\wedge$, а листья T_ϕ — переменными или их отрицаниями. С другой стороны, по любому такому бинарному дереву можно восстановить булеву формулу. Это дает взаимно однозначное представление булевых формул размеченными бинарными деревьями. Если T_ϕ имеет n листьев, то не более n переменных могут встретиться в T_ϕ, поэтому в дальнейшем будем полагать, что все переменные T_ϕ лежат в множестве $\{x_1, \ldots, x_n\}$. *Представление* ϕ состоит из бинарного дерева T_ϕ. Заметим также, что число булевых операций в ϕ равно $n - 1$. Под размером формулы ϕ будем понимать число листьев n.

Например, на рис. 1 представлена формула $(\neg x_1) \wedge (x_2 \vee \neg x_3)$.

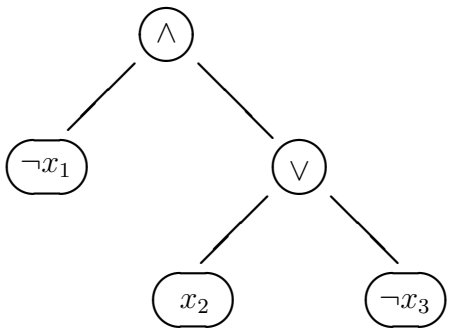

Рис. 1

В дальнейшем будем отождествлять булевы формулы с их представлениями. Обозначим через $\mathcal{F}$ множество всех булевых формул, а через $\mathcal{F}_n$ — множество всех формул в $\mathcal{F}$ размера n. Напомним, что числа Каталана C_n определяются следующим образом:

$$C_n = \frac{1}{n+1}\binom{2n}{n},$$

где $\binom{2n}{n}$ — соответствующий биномиальный коэффициент.

Лемма 1. $|\mathcal{F}_n| = 2^{n-1}(2n)^n C_{n-1}$.

Доказательство. Любая формула из $\mathcal{F}$ размера n есть размеченное бинарное дерево с n листьями и $n-1$ внутренними вершинами. Известно [9], 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 лист может быть помечен одной из n переменных или её отрицанием, поэтому существует $(2n)^n$ таких разметок. Это показывает, что $|\mathcal{F}_n| = 2^{n-1}(2n)^n C_{n-1}$. Лемма доказана. ■

Для любой формулы ϕ определим множество

$$OR(\phi) = \{\phi \vee \psi,\ \psi \text{ — произвольная формула}\}.$$

Лемма 2. Для любой формулы ϕ множество $OR(\phi)$ не строго пренебрежимо. Более того,

$$\frac{|OR(\phi) \cap \mathcal{F}_n|}{|\mathcal{F}_n|} > \frac{1}{(16n)^k}$$

для любого $n > k$, где k — размер формулы ϕ.

Доказательство. Пусть формула ϕ имеет размер k. Тогда для любой формулы $\phi \vee \psi$ из множества $OR(\phi) \cap \mathcal{F}_n$ формула ψ должна иметь размер $n - k$. Кроме того, в этой формуле может участвовать любая из n переменных. Поэтому аналогично тому, как это делалось в доказательстве леммы 1, можно подсчитать

$$|OR(\phi) \cap \mathcal{F}_n| = 2^{n-k-1}(2n)^{n-k} C_{n-k-1}.$$

Отсюда

$$\frac{|OR(\phi) \cap \mathcal{F}_n|}{|\mathcal{F}_n|} = \frac{2^{n-k-1}(2n)^{n-k}C_{n-k-1}}{2^{n-1}(2n)^n C_{n-1}} = \frac{1}{(4n)^k}\frac{C_{n-k-1}}{C_{n-1}} = \frac{1}{(4n)^k}\frac{n}{n-k}\frac{\binom{2(n-k-1)}{n-k-1}}{\binom{2(n-1)}{n-1}} >$$

$$> \frac{1}{(4n)^k}\frac{(n-1)!}{(n-k-1)!}\frac{2(n-k-1)\dots(n-k)}{2(n-1)\dots n} = \frac{1}{(4n)^k}\frac{((n-1)\dots(n-k))^2}{2(n-1)\dots(2n-2k-1)} >$$

$$> \frac{1}{(4n)^k}\Big(\frac{(n-1)\dots(n-k)}{2(n-1)\dots(2n-2k-1)}\Big)^2 > \frac{1}{(4n)^k}\frac{1}{2^{2k}} = \frac{1}{(16n)^k}.$$

Таким образом, имеем $\dfrac{|OR(\Phi)_n|}{|\mathcal{F}_n|} > \dfrac{1}{(16n)^k}$, что и требовалось доказать. ■

3. Основной результат

Теорема 1. Если существует строго генерическое полиномиальное множество булевых формул, на котором проблема общезначимости булевых формул разрешима за полиномиальное время, то существует вероятностный полиномиальный алгоритм, разрешающий эту проблему на всём множестве формул.

Доказательство. Допустим, что существует строго генерическое разрешимое множество формул G, такое, что существует полиномиальный алгоритм $\mathfrak{A}$, определяющий для любой булевой формулы $\phi \in G$, является ли она тождественно истинной. Построим теперь алгоритм $\mathfrak{B}$, определяющий тождественную истинность любой формулы ϕ. На формуле ϕ размера n алгоритм $\mathfrak{B}$ будет работать следующим образом:

1) Проверяет, принадлежит ли ϕ множеству G. Если да, то с помощью алгоритма $\mathfrak{A}$ определяет выполнимость ϕ. Если нет, то переходит к шагу 2.
2) Генерирует случайную формулу ψ размера $n^2 - n$.

3) Проверяет, принадлежат ли формулы $\phi \vee \psi$ и $\phi \vee \neg\psi$ множеству G. Если обе формулы принадлежат G, то с помощью алгоритма $\mathfrak{A}$ определяет их общезначимость и переходит к шагу 4. Если не принадлежат, то выдаёт ответ «НЕ ОБЩЕЗНАЧИМА».
4) Так как обязательно хотя бы одна из формул ψ и $\neg\psi$ не является тождественно истинной, то возможны следующие варианты:
   — Если $\phi \vee \psi$ и $\phi \vee \neg\psi$ тождественно истинны, то ϕ тоже тождественно истинна и алгоритм выдаёт ответ «ОБЩЕЗНАЧИМА».
   — Если хотя бы одна из них не тождественно истинна, то и ϕ тоже не тождественно истинна и алгоритм выдаёт ответ «НЕ ОБЩЕЗНАЧИМА».

   В любом из этих двух случаев алгоритм выдаёт правильный ответ.

Заметим, что алгоритм выдаёт правильный ответ на шагах 1 и 4, а на шаге 3 может выдать неправильный ответ. Нужно доказать, что вероятность того, что ответ выдаётся на шаге 3, меньше $1/2$. Вероятность того, что случайная формула вида $\phi \wedge \psi$ из $OR(\phi)_{n^2}$ не попадёт в G, не больше

$$\frac{|(\mathcal{F} \setminus G)_{n^2}|}{|OR(\phi)_{n^2}|} = \frac{|(\mathcal{F} \setminus G)_{n^2}|}{|\mathcal{F}_{n^2}|} \frac{|\mathcal{F}_{n^2}|}{|OR(\Phi)_{n^2}|}$$

Так как G строго генерическое, то существует константа $\alpha > 0$, такая, что

$$\frac{|(\mathcal{F} \setminus G)_{n^2}|}{|\mathcal{F}_{n^2}|} < \frac{1}{2^{\alpha n^2}}$$

для любого n. С другой стороны, по лемме 2

$$\frac{|\mathcal{F}_{n^2}|}{|OR(\Phi)_{n^2}|} < (16n^2)^n.$$

Поэтому искомая вероятность не больше

$$\frac{(16n^2)^n}{2^{\alpha n^2}} = \frac{2^{4n+2n\log n}}{2^{\alpha n^2}}$$

и при больших n меньше $1/4$. Аналогично делается оценка для формул вида $\phi \vee \neg\psi$. Вероятность же непопадания в G хотя бы одной из формул $\phi \vee \psi$ или $\phi \vee \neg\psi$ не больше $1/4 + 1/4 = 1/2$. Это означает, что вероятность выдачи ответа на шаге 3 меньше $1/2$.

Осталось доказать полиномиальность алгоритма. Для этого нужно за полиномиальное время уметь генерировать случайно и равномерно формулу размера $N = n^2 - n$. Это делается следующим образом:

1) Генерируем некоторую последовательность (далее «слово») из N символов a и $N - 1$ символов p.
2) Делаем такой циклический сдвиг этого слова, чтобы оно начиналось на символ a и заканчивалось на p. Этому слову соответствует обратная польская запись для скобочного выражения от символов a.
3) По слову ищём скобочное выражение следующим образом: пробегаем по всем символам слова, если встречаем символ a, то помещаем его в стек. Если встречаем символ p, то извлекаем два элемента из стека, затем добавляем между ними символ p, заключаем их в скобки и помещаем в стек. Если по ходу процедуры стек окажется пуст, то переходим к шагу 2. Если все пройдёт нормально и мы дойдём до конца слова, и при этом в стеке останется всего один элемент, то искомым скобочным выражением и будет этот элемент. Иначе переходим к шагу 2.

4) Вместо букв p подставляем $\vee$ или $\wedge$ — равновероятно.
5) Каждую букву a в слове заменяем на переменную $x_1, \ldots, x_N$ или её отрицание (какую — выбираем равновероятно).

Корректность этого алгоритма и равномерность генерации формул следует из того, что существует взаимно-однозначное соответствие между обратными польскими записями из N символов a и $N-1$ символов p и бинарными деревьями с N листьями, которые помечены символом a [9].

Итак, в предположении существования полиномиального строго генерического множества, на котором проблема выполнимости булевых формул разрешима за полиномиальное время, построен вероятностный полиномиальный алгоритм, разрешающий эту проблему на всём множестве формул. ■

Непосредственно из теоремы 1 следует

Теорема 2. Если $\mathrm{P} \neq \mathrm{NP}$ и $\mathrm{P} = \mathrm{BPP}$, то не существует строго генерического полиномиального подмножества булевых формул, на котором проблема общезначимости булевых формул разрешима за полиномиальное время.

ЛИТЕРАТУРА

1. *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.
2. *Kapovich I., Miasnikov A., Schupp P., and Shpilrain V.* Average-case complexity for the word and membership problems in group theory // Adv. Math. 2005. V. 190. P. 343–359.
3. *Hamkins J. D. and Miasnikov A. G.* The halting problem is decidable on a set of asymptotic probability one // Notre Dame J. Formal Logic. 2006. V. 47. No. 4. P. 515–524.
4. *Gilman R., Miasnikov A. G., Myasnikov A. D., and Ushakov A.* Report on generic case complexity // Herald of Omsk University. 2007. Special Issue. P. 103–110.
5. *Cook S. A.* The complexity of theorem proving procedures // Proc. 3d Annual ACM Symposium on Theory of Computing. N. Y., USA, 1971. P. 151–158.
6. *Impagliazzo R. and Wigderson A.* $\mathrm{P} = \mathrm{BPP}$ unless E has subexponential circuits: Derandomizing the XOR Lemma // Proc. 29th STOC. El Paso: ACM, 1997. P. 220–229.
7. *Myasnikov A. and Rybalov A.* Generic complexity of undecidable problems // J. Symbolic Logic. 2008. V. 73. No. 2. P. 656–673.
8. *Rybalov A.* Generic complexity of presburger arithmetic // Theory Comput. Systems. 2010. V. 46. No. 1. P. 2–8.
9. *Кнут Д.* Искусство программирования. М.: Вильямс, 2010. 720 с.

REFERENCES

1. *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.
2. *Kapovich I., Miasnikov A., Schupp P., and Shpilrain V.* Average-case complexity for the word and membership problems in group theory. Adv. Math., 2005, vol. 190, pp. 343–359.
3. *Hamkins J. D. and Miasnikov A. G.* The halting problem is decidable on a set of asymptotic probability one. Notre Dame J. Formal Logic, 2006, vol. 47, no. 4, pp. 515–524.
4. *Gilman R., Miasnikov A. G., Myasnikov A. D., and Ushakov A.* Report on generic case complexity. Herald of Omsk University, 2007, Special Issue, pp. 103–110.
5. *Cook S. A.* The complexity of theorem proving procedures. Proc. 3d Annual ACM Symposium on Theory of Computing, N. Y., USA, 1971, pp. 151–158.

6. *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.
7. *Myasnikov A. and Rybalov A.* Generic complexity of undecidable problems. J. Symbolic Logic, 2008, vol. 73, no. 2, pp. 656–673.
8. *Rybalov A.* Generic complexity of presburger arithmetic. Theory Comput. Systems, 2010, vol. 46, no. 1, pp. 2–8.
9. *Knuth D. E.* The Art of Computer Programming. Reading, Massachusetts, Addison-Wesley, 1997.

СВЕДЕНИЯ ОБ АВТОРАХ

БЫЛИНА Роман Анатольевич — г. Москва. E-mail: **bopobey@rambler.ru**

ДАВИДОВСКИЙ Максим Владимирович — ассистент кафедры информационных технологий Запорожского национального университета, г. Запорожье.
E-mail: **m.davidovsky@gmail.com**

ДЕНИСОВ Олег Викторович — кандидат физико-математических наук, доцент, ООО «Центр сертификационных исследований», г. Москва.
E-mail: **denisovOleg@yandex.ru**

ЗУБОВ Анатолий Юрьевич — кандидат физико-математических наук, доцент, старший научный сотрудник Московского государственного университета им. М. В. Ломоносова, г. Москва. E-mail: **zubovanatoly@yandex.ru**

КУРАПОВ Сергей Всеволодович — кандидат физико-математических наук, доцент, доцент кафедры математического моделирования Запорожского национального университета, г. Запорожье. E-mail: **lilili5050@rambler.ru**

ЛУКЬЯНОВА Наталья Александровна — старший преподаватель Института математики и фундаментальной информатики Сибирского федерального университета, г. Красноярск. E-mail: **nata00sfu@gmail.com**

МОНАРЁВ Виктор Александрович — кандидат физико-математических наук, научный сотрудник Института вычислительных технологий СО РАН, г. Новосибирск. E-mail: **viktor.monarev@gmail.com**

НОВОСЕЛОВ Семен Александрович — аспирант Балтийского федерального университета им. И. Канта, г. Калининград. E-mail: **snovoselov@kantiana.ru**

ПЕСТУНОВ Андрей Игоревич — кандидат физико-математических наук, доцент Новосибирского государственного университета экономики и управления, г. Новосибирск. E-mail: **pestunov@gmail.com**

РЫБАЛОВ Александр Николаевич — кандидат физико-математических наук, старший научный сотрудник Института математики им. С. Л. Соболева СО РАН, г. Новосибирск, Россия. E-mail: **alexander.rybalov@gmail.com**

САЛИЙ Вячеслав Николаевич — кандидат физико-математических наук, профессор, заведующий кафедрой теоретических основ компьютерной безопасности и криптографии Саратовского государственного университета им. Н. Г. Чернышевского, г. Саратов. E-mail: **SaliiVN@info.sgu.ru**

СЕМЕНОВА Дарья Владиславовна — кандидат физико-математических наук, доцент, доцент Института математики и фундаментальной информатики Сибирского федерального университета, г. Красноярск. E-mail: **dariasdv@gmail.com**

СОШИН Данил Андреевич — сотрудник ФГУП «НИИ «Квант», г. Москва.
E-mail: **danil_re@list.ru**

Журнал «Прикладная дискретная математика» включен в перечень ВАК рецензируемых российских журналов, в которых должны быть опубликованы основные результаты диссертаций, представляемых на соискание учёной степени кандидата и доктора наук, в базу данных Web of Science (Russian Science Citation Index — RSCI), а также в перечень журналов, рекомендованных УМО в области информационной безопасности РФ в качестве учебной литературы по специальности «Компьютерная безопасность».

Журнал «Прикладная дискретная математика» распространяется по подписке; его подписной индекс 38696 в объединённом каталоге «Пресса России». Полнотекстовые электронные версии вышедших номеров журнала доступны на его сайте journals.tsu.ru/pdm и на Общероссийском математическом портале www.mathnet.ru. На сайте журнала можно найти также и правила подготовки рукописей статей в журнал.

Тематика публикаций журнала:

- *Теоретические основы прикладной дискретной математики*
- *Математические методы криптографии*
- *Математические методы стеганографии*
- *Математические основы компьютерной безопасности*
- *Математические основы надёжности вычислительных и управляющих систем*
- *Прикладная теория кодирования*
- *Прикладная теория автоматов*
- *Прикладная теория графов*
- *Логическое проектирование дискретных автоматов*
- *Математические основы информатики и программирования*
- *Вычислительные методы в дискретной математике*
- *Дискретные модели реальных процессов*
- *Математические основы интеллектуальных систем*
- *Исторические очерки по дискретной математике и её приложениям*