МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ

ПРИКЛАДНАЯ ДИСКРЕТНАЯ МАТЕМАТИКА

Научный журнал

2015 № 2(28)

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 15.06.2015.
Формат $60 \times 84\frac{1}{8}$. Усл. п. л. 12,6. Уч.-изд. л. 14. Тираж 300 экз. Заказ № 1103.

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

СОДЕРЖАНИЕ

CONTENTS

ТЕОРЕТИЧЕСКИЕ ОСНОВЫ ПРИКЛАДНОЙ ДИСКРЕТНОЙ МАТЕМАТИКИ

УДК 519.233.33+519.244.4

СТАТИСТИЧЕСКИЕ МЕТОДЫ ПОИСКА НАБОРА КООРДИНАТ, НА КОТОРОМ СЛУЧАЙНЫЙ ВЕКТОР ИМЕЕТ ЗАПРЕТЫ

О. В. Денисов

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

Наблюдается стационарная последовательность случайных векторов длины L, имеющих распределение случайного вектора ξ; координаты векторов принимают значения в конечном множестве. Рассматривается гипотеза о существовании некоторого множества номеров координат $\Theta \subset \{1, \ldots, L\}$, такого, что подвектор ξ_Θ (проекция ξ на координаты с номерами из Θ) распределён как заданный случайный вектор η, распределение которого имеет запреты. Строится критерий согласия на основе анализа запретов эмпирического распределения. Когда априори известно, что гипотеза выполнена, предлагаются три алгоритма поиска части Θ, работающие при разной доле информации о распределении случайного вектора η.

Ключевые слова: *статистический критерий, запреты распределений.*

DOI 10.17223/20710410/28/1

STATISTICAL METHODS OF SEARCH FOR COORDINATE SET ON WHICH A RANDOM VECTOR HAS BANS

O. V. Denisov

Certification Research Center, Moscow, Russia

E-mail: denisovOleg@yandex.ru

A stationary sequence of random vectors of length L with the distribution of a random vector ξ is observed. Coordinates of vectors in it take values in a finite set. The following hypothesis is considered: there is a set $\Theta \subset \{1, \ldots, L\}$ such that the subvector ξ_Θ (being the projection of ξ onto coordinates with numbers in Θ) has the distribution of a given random vector η with the distribution having bans. A concordance criterion is constructed by the analysis of an empirical distribution bans. In the case of the hypothesis validity (a priori), three algorithms to search for a part of Θ are proposed. They work under various portions of the information about the random vector η distribution.

Keywords: *statistical test, bans of distributions.*

Введение

Понятие запретов дискретного вероятностного распределения введено в рабо тах [1, 2]; в [2] предложен статистический метод поиска минимальных запретов. В частности, такие методы могут иметь приложения в стеганографии [1, пример 3].

В данной работе эти подходы применяются при решении более конкретно поставленной задачи проверки статистической гипотезы о наличии некоторого частного распределения заданного вида у наблюдаемого многомерного распределения. Для этого развивается понятийный аппарат, связанный с запретами. Главными новыми понятиями являются «графы простых запретов» и «устойчивость координаты случайного вектора к запретам». На их основе построен критерий согласия с гипотезой и три алгоритма поиска некоторых номеров координат из тех, на которых сосредоточено частное распределение; получены оценки вероятностей ошибок критерия и алгоритмов.

Заметим, что ранее запреты распределения выхода простейших неавтономных автоматов исследовались в связи с криптографическими приложениями [3, 4]. При этом определение запрета давалось несколько иначе — в терминах решений систем автоматных уравнений. Для двоичного регистра сдвига понятие запрета введено С. Н. Сумароковым в 1968 г.

Перейдём к строгой постановке задачи. Пусть X — произвольное конечное множество мощности k. Наблюдается отрезок стационарной последовательности случайных векторов длины L

$$\mathbf{x}(t) = (x_1(t), \dots, x_L(t)) \in X^L, \quad 1 \leqslant t \leqslant N,$$

имеющих распределение случайного вектора ξ: $\mathbf{x}(t) \sim \xi$, $1 \leqslant t \leqslant N$, N — длина отрезка (число наблюдений).

Рассматривается задача проверки сложной статистической гипотезы $H(\eta)$ о наличии *особенного* множества номеров координат

$$\Theta = \{\theta_1, \dots, \theta_M\} \subset \{1, \dots, L\},$$

такого, что соответствующий подвектор на этих координатах ξ_Θ имеет распределение случайного вектора

$$\eta = (\eta_1, \dots, \eta_M),$$

известной длины $M < L$, которое полностью или частично известно. Далее предполагается, что оно имеет запреты; строгое определение запрета распределения приведено ниже.

Через $\mathbf{x}_I = (x_{i_1}, \dots, x_{i_r})$ здесь и далее обозначаем подвектор вектора $\mathbf{x}$, состоящий из координат с номерами из множества $I = \{i_1, \dots, i_r\}$. Для произвольного множества A обозначим через 2^A множество всех подмножеств множества A; $A^{(r)}$ — множество всех r-элементных подмножеств A; $\bar{A}$ — дополнение к множеству A.

Итак, гипотеза $H(\eta)$ формулируется как гипотеза о том, что ξ имеет соответствующее частное распределение:

$$H(\eta) = \Big\{ \exists\, \Theta \subset \{1, \dots, L\}^{(M)} : \ \xi_{\boldsymbol{\Theta}} \sim \eta \Big\}.$$

Если априори известно, что предположение $H(\eta)$ выполнено, то встаёт вопрос об оценке множества Θ как параметра распределения.

Далее в п. 1 введён ряд новых понятий, связанных с носителями и запретами частных распределений, попутно доказывая некоторые их свойства. В п. 2 построен статистический критерий проверки гипотезы $H(\eta)$ и при некоторых предположениях получены оценки его ошибок. Затем предлагаются три алгоритма поиска части множества Θ, работающие при разной доле информации о распределении η. При полной информации о η предлагается последовательный алгоритм 1; он имеет нулевую вероятность ошибки. Алгоритмы 2 и 3 строятся при отсутствии информации о распределении η и работают на фиксированном объёме материала.

В п. 3 вводятся ограничения на распределение ξ, упрощающие построение критерия, расчёт параметров критерия и алгоритмов. Для схемы независимых наблюдений получены верхние оценки для числа наблюдений N, при котором вероятности ошибок не превосходят заданной величины.

Согласно [2, с.57], данные методы могут быть применены для «...статистического выявления скрытых каналов, в которых вставки осуществляются с помощью некоторых функциональных соотношений». У нас это соответствует задаче поиска распределения вида «вектор аргументов и вектор-функция от него» при фиксированной $\mathbf{f}(\mathbf{x}) : X^n \to X^m$, где вектор аргументов $\mathbf{x}$ не имеет запретов. В простейшем случае, когда $\mathbf{x} = (x_1, \ldots, x_n)$ распределён равномерно, случайный вектор обозначаем через

$$\mathbf{xf} = (x_1, \ldots, x_n, f_1(\mathbf{x}), \ldots, f_m(\mathbf{x})), \quad \mathbf{x} \sim U(X^n).$$

1. Основные понятия

Введём сначала теоретические характеристики распределения ξ, связанные с запретами, а затем эмпирические.

1.1. Запреты распределения случайного вектора и графы запретов

Случайный вектор ξ имеет дискретное распределение, поэтому далее без ограничения общности считаем, что *носитель распределения* ξ

$$\mathrm{Supp}(\xi) = \left\{\mathbf{x} \in X^L : \mathbf{P}\{\xi = \mathbf{x}\} > 0\right\}$$

совпадает с множеством значений случайного вектора: $\mathrm{Supp}(\xi) = \xi(\Omega)$, то есть ξ принимает все свои значения с положительной вероятностью. Разрабатываемые методы основаны на анализе только наблюдаемых носителей частных распределений ξ.

Введём ряд определений и обозначений. Вектор $\mathbf{a} \in X^L$ называется *запретом размерности L распределения* ξ, если $\mathbf{P}\{\xi = \mathbf{a}\} = 0$. При этом $\mathbf{a}$ называется *простым запретом размерности* L, если

$$\mathbf{P}\left\{\xi_{\{1,\ldots,L\}\setminus\{i\}} = \mathbf{a}_{\{1,\ldots,L\}\setminus\{i\}}\right\} > 0,\ 1 \leqslant i \leqslant L,$$

то есть любой его собственный подвектор не является запретом распределения любого подвектора ξ.

Множество всех запретов размерности L распределения ξ обозначим через $\mathcal{Z}(\xi)$, а простых запретов размерности L — через $\mathcal{Z}_s(\xi)$. Очевидно, что

$$\mathcal{Z}_s(\xi) \subset \mathcal{Z}(\xi) = X^L \setminus \mathrm{Supp}(\xi).$$

Принадлежность вектора к запретам может определяться лишь значениями некоторых его координат. Чтобы далее выявлять такие наборы координат, рассмотрим запреты частных распределений.

Вектор $\mathbf{a} \in X^r$ будем называть *запретом размерности r распределения ξ на множестве (номеров координат)* $J \in \{1, \dots, L\}^{(r)}$, если $\mathbf{a} \in \mathcal{Z}(\xi_J)$, $1 \leqslant r \leqslant L$. При этом $\mathbf{a}$ будем называть *простым запретом распределения ξ (размерности r)*, если $\mathbf{a} \in \mathcal{Z}_s(\xi_J)$. Это равносильно тому, что для $J = \{j_1, \dots, j_r\}$ выполнено

$$\mathbf{P}\{\xi_J = \mathbf{a}\} = 0, \qquad \mathbf{P}\left\{\xi_{J\setminus\{j_s\}} = \mathbf{a}_{\{1,\dots,r\}\setminus\{s\}}\right\} > 0,\ 1 \leqslant s \leqslant r.$$

Кратчайшими запретами распределения ξ будем называть запреты наименьшей размерности. Эту размерность назовём *устойчивостью к запретам* распределения ξ и обозначим $\mathbf{z}_{\min}(\xi)$.

Очевидно, что кратчайшие запреты всегда являются простыми. Из определений также следует, что свойство «быть простым запретом» сохраняется при расширении случайного вектора: для $|J| = r$, $\mathbf{a} \in X^r$ выполнено

$$\begin{gathered}\mathbf{a} \in \mathcal{Z}_s(\xi_J) \text{ тогда и только тогда, когда } \mathbf{a} \text{ является простым запретом} \\ \text{размерности } r \text{ распределения } \xi_I \text{ для всех } I \supset J.\end{gathered} \tag{1}$$

Заметим, что каждый запрет размерности r на наборе J влечёт появление k запретов размерности $r+1$ на наборах вида $J \cup \{i\}$, $i \notin J$. Такие запреты размерности $r+1$ не несут новой информации о множестве $\mathrm{Supp}(\xi)$ при знании всех запретов размерности r, и в этом смысле интересны лишь простые запреты.

Теперь введём основные инструменты анализа.

Определение 1. *Графом запретов размерности r распределения ξ* назовём r-однородный гиперграф (кратко r-граф) на вершинах с номерами из $\{1, \dots, L\}$

$$\mathcal{G}(r, \xi) = \left\{J \subset \{1, \dots, L\}^{(r)} : \mathcal{Z}_s(\xi_J) \neq \varnothing\right\}.$$

Его рёбрами являются все наборы, на которых распределение ξ имеет простые запреты размерности r. Граф

$$\mathcal{G}_{\min}(\xi) = \mathcal{G}(\mathbf{z}_{\min}(\xi), \xi)$$

назовём *графом кратчайших запретов*. Его можно определить также как первый непустой r-граф в цепочке $\mathcal{G}(1, \xi)$, $\mathcal{G}(2, \xi)$, …, $\mathcal{G}(L, \xi)$.

Везде далее будем считать, что $\mathcal{Z}(\xi) \neq \varnothing$ (в противном случае анализ не даст никакой информации о распределении ξ), и тогда в этой цепочке есть хотя бы один непустой r-граф.

Обозначим также через

$$\mathcal{G}(\xi) = \bigsqcup_{1 \leqslant r \leqslant L} \mathcal{G}(r, \xi) \subset 2^{\{1,\dots,L\}}$$

гиперграф (неоднородный в общем случае), состоящий из наборов всех простых запретов.

Для произвольного гиперграфа $\mathcal{G} \subset 2^{\{1,\dots,L\}}$ через

$$\mathbf{V}(\mathcal{G}) = \bigcup_{J \in \mathcal{G}} J, \quad \mathbf{v}(\mathcal{G}) := |\mathbf{V}(\mathcal{G})|$$

обозначим соответственно множество тех вершин гиперграфа $\mathcal{G}$, которые покрыты хотя бы одним его ребром, и число таких вершин.

Определение 2. Будем говорить, что r-графы $\mathcal{G}$ и $\mathcal{G}'$ *изоморфны* ($\mathcal{G} \cong \mathcal{G}'$), если существует биекция $\phi : \mathbf{V}(\mathcal{G}) \to \mathbf{V}(\mathcal{G}')$, такая, что

$$J \in \mathcal{G} \Longleftrightarrow \phi(J) \in \mathcal{G}.$$

Это означает, что $\mathcal{G}'$ может быть получен из $\mathcal{G}$ взаимно однозначным изменением номеров вершин после удаления из $\mathcal{G}$ и $\mathcal{G}'$ вершин, не лежащих ни в одном ребре. Необходимым условием изоморфности, очевидно, является равенство $\mathbf{v}(\mathcal{G}) = \mathbf{v}(\mathcal{G}')$.

Для r-графа $\mathcal{G}$ и множества $I \subset \{1, \ldots, L\}$, $1 \leqslant r \leqslant |I|$, через

$$\mathcal{G}_I = \mathcal{G} \cap I^{(r)}$$

обозначим ограничение $\mathcal{G}$ на множество вершин I, то есть соответствующий r-подграф $\mathcal{G}$. Из условия (1) имеем равенство $\mathcal{G}(r, \xi)_I = \mathcal{G}(r, \xi_I)$. Отсюда следует, что при гипотезе $H(\eta)$ имеем

$$\mathcal{G}(r, \xi)_\Theta = \mathcal{G}(r, \xi_\Theta) \cong \mathcal{G}(r, \eta), \quad 1 \leqslant r \leqslant M.$$

На этом свойстве основан критерий согласия с гипотезой $H(\eta)$ и статистические алгоритмы поиска частей множества Θ, в том числе его части:

$$\Theta_{\min} = \mathbf{V}(\mathcal{G}_{\min}(\xi_\Theta)) \subset \Theta.$$

Так как $\mathcal{G}_{\min}(\xi_\Theta) \cong \mathcal{G}_{\min}(\eta)$, то $|\Theta_{\min}| = \mathbf{v}(\mathcal{G}_{\min}(\eta))$.

1.2. Эмпирические характеристики распределения

Заметим, что введённые выше характеристики инвариантны относительно любого изменения распределения вероятностей ξ, при котором сохраняется носитель распределения. Поэтому их определения можно дать без использования понятия вероятности, оперируя лишь множеством $\mathrm{Supp}(\xi)$.

Продемонстрируем этот путь при определении аналогичных эмпирических характеристик распределения ξ. Обозначим через

$$\mathcal{X} = \mathcal{X}(N) \subset \mathrm{Supp}(\xi) \subset X^L$$

множество всех различных векторов среди наблюдений $\mathbf{x}(1), \ldots, \mathbf{x}(N)$. Элементы множества $\mathcal{Z}(\mathcal{X}) = X^r \backslash \mathcal{X}$ назовём *эмпирическими запретами* (*запретами множества* $\mathcal{X}$) *размерности* L.

Введём операцию ограничения $\mathcal{X}$ на множество координат с номерами из $I \in \{1, \ldots, L\}^{(r)}$ (*проекции* $\mathcal{X}$ на I): $\mathcal{X}_I = \{\mathbf{a}_I : \mathbf{a} \in \mathcal{X}\} \subset X^r$.

Вектор $\mathbf{a} \in \mathcal{Z}(\mathcal{X})$ называется *простым эмпирическим запретом размерности* L, если $\mathbf{a}_I \in \mathcal{X}_I$ для всех $\varnothing \neq I \subset \{1, \ldots, L\}$. Множество всех простых эмпирических запретов размерности L обозначим через $\mathcal{Z}_s(\mathcal{X})$.

Аналогично запретам распределения, для $J \in \{1, \ldots, L\}^{(r)}$ элементы множеств $\mathcal{Z}(\mathcal{X}_J)$ и $\mathcal{Z}_s(\mathcal{X}_J)$ называем *r-мерными эмпирическими запретами* (соответственно *простыми r-мерными эмпирическими запретами*) на множестве J; определяем *r-графы эмпирических запретов* $\mathcal{G}(r, \mathcal{X})$, $1 \leqslant r \leqslant L$, и их объединение — гиперграф $\mathcal{G}(\mathcal{X})$.

Заметим, что все эти понятия можно было ввести сразу на базе предыдущих определений, формально интерпретируя параметр $\mathcal{X}$ как обозначение для некоторого произвольного распределения с носителем $\mathcal{X}$.

Множества $\mathcal{X}$ и $\mathcal{X}_I$ являются статистическими оценками носителей $\mathrm{Supp}(\xi)$ и $\mathrm{Supp}(\xi_{\mathbf{I}})$, всегда лежащими в них, и поэтому всегда $\mathcal{Z}(\mathcal{X}_I) \supset \mathcal{Z}(\xi_I)$, $I \subset \{1,\dots,L\}$.

Легко видеть, что

$$\begin{gathered}\text{всегда } \mathbf{z}_{\min}(\mathcal{X}) \leqslant \mathbf{z}_{\min}(\xi), \text{ и если здесь достигается равенство,} \\ \text{то } \mathcal{G}_{\min}(\mathcal{X}) \supset \mathcal{G}_{\min}(\xi).\end{gathered} \tag{2}$$

Для множеств простых запретов включение $\mathcal{G}(r,\mathcal{X}(N)) \supset \mathcal{G}(r,\xi)$ в общем случае неверно, что показывает следующий

Пример 1. Пусть $X = \mathbb{Z}_2$, $L = 2$, $\mathcal{Z}(\xi) = \{(1,0)\}$. Тогда одномерных запретов ξ не имеет, $\mathcal{G}(1,\xi) = \varnothing$, а имеющийся запрет размерности 2 является простым, и граф $\mathcal{G}_{\min}(\xi) = \mathcal{G}(2,\xi)$ состоит из единственного ребра $\{1,2\}$.

Далее, пусть $\mathcal{X} = \{(0,0)\}$. Тогда эмпирическими запретами размерности 2 являются все ненулевые векторы, вектор (1) является простым запретом на множествах $\{1\}$ и $\{2\}$, $\mathcal{G}(1,\mathcal{X}) = \{\{1\},\{2\}\} \supset \mathcal{G}(1,\xi)$. Отсюда также следует, что любой ненулевой вектор длины 2 не является простым эмпирическим запретом, и поэтому $\mathcal{Z}_s(\mathcal{X}) = \varnothing$, $\mathcal{G}(2,\mathcal{X}) = \varnothing$.

Этот пример также показывает существенность условия о равенстве в (2).

1.3. Предположение о вероятностной модели

Далее будем считать, что для любых $0 < \alpha < 1$ и $1 \leqslant R \leqslant L$ определена функция $N_1 = N_1(R,\xi,\alpha)$, такая, что

$$\mathbf{P}\left\{\forall J \in \{1,\dots,L\}^{(R)}\ \ (\mathcal{X}(N)_J = \mathrm{Supp}(\xi_J))\right\} \geqslant 1-\alpha \tag{3}$$

при $N \geqslant N_1(R,\xi,\alpha)$. Это означает, что при всех достаточно больших N с вероятностью не менее заданной все R-мерные проекции выборки принимают все возможные для них значения. Заметим, что функция N_1 зависит не только от распределения ξ в фиксированный момент времени, но и от распределения $\{\mathbf{x}(t)\}_{t\geqslant 1}$ всей последовательности состояний. Для упрощения обозначений эту зависимость будем иметь в виду, не указывая явно.

Очевидно, что при $N \geqslant N_1(R,\xi,\alpha)$ справедливы следующие оценки:

$$\begin{gathered}\mathbf{P}\left\{\forall I, 1 \leqslant |I| \leqslant R\ \ (\mathcal{Z}(\mathcal{X}(N)_I) = \mathcal{Z}(\xi_I), \mathcal{Z}_s(\mathcal{X}(N)_I) = \mathcal{Z}_s(\xi_I))\right\} \geqslant 1-\alpha; \\ \mathbf{P}\left\{\forall\, r \in \{1,\dots,R\}\ (\mathcal{G}(r,\mathcal{X}(N)) = \mathcal{G}(r,\xi))\right\} \geqslant 1-\alpha.\end{gathered} \tag{4}$$

1.4. Гипотеза о наличии функциональной вставки

Важным частным случаем гипотезы $H(\eta)$ является случай, когда η имеет распределение вида «вектор аргументов и функции от него», где вектор аргументов не имеет запретов. В частности, это выполнено для $\eta \sim \mathbf{xf}$. Тогда при фиксированной $\mathbf{f} : X^n \to X^m$ сложная статистическая гипотеза $H(\eta)$ строго формулируется так:

$$\exists \Theta = Inp \sqcup Out \in \{1,\dots,L\}^{(n+m)}\ \Big(\xi(\Omega)_{Inp} = X^n,\ \forall \omega \in \Omega\ \left(\xi(\omega)_{Out} = \mathbf{f}\left(\xi(\omega)_{Inp}\right)\right)\Big). \tag{5}$$

Здесь множества Inp, Out образуют разбиение множества Θ; первое может являться множеством номеров аргументов, а второе — множеством номеров значений функции.

Заметим, что гипотеза $H(\eta)$ является существенным обобщением (5) и включает также случаи, когда на координатах с номерами особенного множества реализована не вектор-функция, а нечто более сложное.

Рассмотрим пример, демонстрирующий идею использования графов запретов при поиске частных распределений вида $\mathbf{xf}$.

Пример 2. Пусть $X = \mathbb{Z}_2$, $L \geqslant 5$ и по наблюдениям над реализациями случайного вектора $\xi = (\xi_1, \xi_2, \ldots, \xi_{L-2}, \xi_{L-1} = \xi_1\xi_2, \xi_L = \xi_1 \oplus \xi_2 \oplus \xi_3)$, $(\xi_1, \ldots, \xi_{L-2}) \sim U(\mathbb{Z}_2^{L-2})$, требуется найти функциональную вставку-конъюнкцию.

Решение. Здесь моделью частного распределения на особенном множестве координат является вероятностная схема $\eta = \mathbf{xf} = (x_1, x_2, x_1x_2)$, $(x_1, x_2) \sim U(\mathbb{Z}_2^2)$ вида $\mathbf{xf}$. Для неё $\mathcal{G}(1, \mathbf{xf}) = \varnothing$ и $\mathcal{G}_{\min}(\mathbf{xf}) = \mathcal{G}(2, \mathbf{xf}) = \big\{\{1, 3\}, \{2, 3\}\big\}$.

Одномерные распределения ξ также не имеют запретов, двумерные распределения имеют запрет (0,1) на наборах координат из $\mathcal{G}(2, \xi) = \big\{\{1, L-1\}, \{2, L-1\}\big\} \cong \mathcal{G}(2, \mathbf{xf})$. Тогда при $N \geqslant N_1(r, \xi, \alpha)$, $r = 2$ с вероятностью не менее $1-\alpha$, согласно (4), выполнено $\mathcal{G}(2, \mathcal{X}) = \mathcal{G}(2, \xi)$. Это позволяет сделать вывод, что вставка-конъюнкция возможна только на координатах с номерами из $\Theta^* = \{1, 2, L-1\}$, причём $x_{L-1} = f(x_1, x_2)$.

Устойчивость к запретам координат случайного вектора

Заметим, что если в примере 2 положить $\xi_L = \xi_1 \oplus \xi_2$, то ξ на $J = \{L-1, L\}$ будет иметь простой запрет (1,1), поскольку $\xi_1\xi_2 = 1 \Rightarrow \xi_1 = \xi_2 = 1 \Rightarrow \xi_1 \oplus \xi_2 = 0$. Тогда граф $\mathcal{G}(2, \mathcal{X}(N))$ при всех N будет содержать ребро $\{L-1, L\}$. Это нарушит изоморфизм $\mathcal{G}(2, \mathcal{X}(N))$ и $\mathcal{G}(2, \mathbf{xf})$ при всех N, и метод определения Θ в указанном виде будет неприменим.

Чтобы избежать таких сложностей, введём ограничение на распределение ξ: все кратчайшие запреты распределения ξ образованы координатами с номерами из Θ. Оно соответствует естественному предположению о том, что сначала осуществляется поиск подвектора, наиболее уязвимого для нашего метода. Далее всюду будем считать предположение выполненным. Оно записывается проще с помощью следующих определений.

Устойчивостью к запретам i-й координаты случайного вектора ξ назовём величину

$$\mathbf{z}(i, \xi) = \min\{|J| : J \ni i, \mathcal{Z}_s(\xi_J) \neq \varnothing\}.$$

Она равна наименьшему значению r, при котором $i \in \mathcal{G}(r, \xi)$, то есть номер i участвует в простом запрете размерности r. Если i не участвует ни в одном простом запрете, то по определению считаем $\mathbf{z}(i, \xi) = \infty$.

Очевидно, что

$$\mathbf{z}_{\min}(\xi) = \min_{1 \leqslant i \leqslant L} \mathbf{z}(i, \xi).$$

Легко видеть, что в примере 2 справедлива импликация $(\xi_3\xi_{L-1} = 1 \Rightarrow \xi_L = 1)$, откуда следует, что $\{3, L-1, L\}$ — ребро графа $\mathcal{G}(3, \xi)$. Поэтому с учётом далее доказываемой теоремы 5, п. 3, имеем

$$\mathbf{z}(i, \xi) = \begin{cases} 2 & \text{при } i \in \{1, 2, L-1\}, \\ 3 & \text{при } i \in \{3, L\}, \\ \infty & \text{при } 4 \leqslant i \leqslant L-2. \end{cases} \tag{6}$$

Рассмотрим также максимальную устойчивость координат к запретам

$$\mathbf{z}_{\max}(\xi) = \max_{1 \leqslant i \leqslant L} \mathbf{z}(i, \xi).$$

Теперь сформулированное ограничение записывается так: при условии $H(\eta)$ выполнено условие

$$\forall i \notin \Theta \quad (\mathbf{z}(i, \xi) > \mathbf{z}_{\min}(\xi_\Theta)).$$

Оно эквивалентно тому, что $\mathcal{G}_{\min}(\xi)$ не содержит рёбер с номерами из $\bar{\Theta}$, то есть равенству

$$\mathcal{G}_{\min}(\xi) = \mathcal{G}_{\min}(\xi_\Theta). \tag{7}$$

Согласно (6), в примере 2 это условие выполнено.

2. Проверка гипотезы и поиск особенного множества

На основе введённых графов запретов сначала построим критерий согласия с гипотезой $H(\eta)$. В случае априорной справедливости $H(\eta)$ предложен последовательный алгоритм 1 определения $\Theta_{\min}$.

2.1. Критерий на основе графа кратчайших запретов

Предлагается следующий критерий согласия с гипотезой $H(\eta)$:

$$\Big(\mathcal{G}_{\min}(\mathcal{X}(N)) \cong \mathcal{G}_{\min}(\eta)\Big) \quad \Longrightarrow \text{ принимаем гипотезу } H(\eta). \tag{8}$$

Докажем теорему о вероятностях ошибок критерия.

Теорема 1. Пусть $\mathbf{z} = \mathbf{z}_{\min}(\eta) < \infty$. Тогда:

1. При любом $N \geqslant 1$ критерий (8) с вероятностью 1 отклоняет все альтернативы ξ, у которых $\mathbf{z}_{\min}(\xi) < \mathbf{z}$ или $(\mathbf{z}_{\min}(\xi) = \mathbf{z},\ |\mathcal{G}_{\min}(\xi)| > |\mathcal{G}_{\min}(\eta)|)$.
2. При $N \geqslant N_1(\mathbf{z}, \xi, \alpha)$ и альтернативе ξ, такой, что $\mathbf{z}_{\min}(\xi) > \mathbf{z}$, вероятность ошибки критерия не превосходит α.
3. Если выполнено ограничение (7) и $N \geqslant N_1(\mathbf{z}, \xi, \alpha)$, то вероятность ошибки критерия при гипотезе $H(\eta)$ не превосходит α.

Доказательство.

1. Обозначим через A условие-предпосылку в (8), а через $r = \mathbf{z}_{\min}(\mathcal{X})$ — число вершин в рёбрах графа $\mathcal{G}_{\min}(\mathcal{X})$.

Если $r < \mathbf{z}$, то A не выполнено и гипотеза отвергается. С учётом (2), при альтернативе первого вида из п. 1 теоремы имеем

$$r \leqslant \mathbf{z}_{\min}(\xi) < \mathbf{z}.$$

Если $r = \mathbf{z}$ и выполнена альтернатива второго вида, то, согласно (2), граф $\mathcal{G}_{\min}(\mathcal{X})$ содержит граф $\mathcal{G}(\mathbf{z}, \xi)$ и, следовательно, число его рёбер больше числа рёбер $\mathcal{G}(\mathbf{z}, \xi)$. Поэтому здесь также невозможно условие A. Пункт 1 доказан.

2. При этой альтернативе $\mathcal{G}(\mathbf{z}, \xi) = \varnothing$. Поэтому, согласно (4), для $N \geqslant N_1(\mathbf{z}, \xi, \alpha)$ с вероятностью не меньше $1-\alpha$ происходит событие $\mathcal{G}(\mathbf{z}, \mathcal{X}) = \varnothing$, которое несовместно с событием A.

3. Из условий $H(\eta)$ и (7) имеем $\mathcal{G}_{\min}(\eta) \cong \mathcal{G}_{\min}(\xi_\Theta) = \mathcal{G}_{\min}(\xi)$. При $N \geqslant N_1(\mathbf{z}, \xi, \alpha)$, согласно (4), с вероятностью не меньше $1 - \alpha$ происходит событие $\mathcal{G}_{\min}(\xi) = \mathcal{G}_{\min}(\mathcal{X})$, что с учётом предыдущего равенства влечёт событие A. ■

Замечание 1. Особенностью критерия является то, что при его применении не накладывается никаких ограничений на распределение ξ, кроме (7). Но от распределения $\{\mathbf{x}(t)\}_{t\geqslant 1}$ и, в частности, от распределения ξ зависит объём материала $N_1(\cdot)$, достаточный для гарантированной верхней оценки вероятности ошибки критерия.

Замечание 2. Основной вклад в сложность проверки условия (8) может вносить построение графа $\mathcal{G}(\mathbf{z}, \mathcal{X}(N))$, $\mathbf{z} = \mathbf{z}_{\min}(\eta)$. Этот граф можно строить с помощью $\binom{L}{\mathbf{z}}$ битовых массивов длины $k^{\mathbf{z}}$, соответствующих всем $J \in \{1, \ldots, L\}^{(\mathbf{z})}$ и первоначально инициализированных нулями. Проходя по всем наблюдаемым векторам $\mathbf{x}(t)$,

$1 \leqslant t \leqslant N$, будем записывать единицы по адресам $\mathbf{x}(t)_J$ для всех J. После этого нулевые элементы в массиве позволят определить запреты и простые запреты размерности $\mathbf{z}$ эмпирического распределения.

Временная и емкостная сложности такого алгоритма оцениваются величинами порядка $\binom{L}{\mathbf{z}}N$, $\binom{L}{\mathbf{z}}k^{\mathbf{z}}$ соответственно. При больших L они быстро растут с ростом $\mathbf{z}$. Поэтому в критерии согласия для достижения малой временной сложности и для простоты критерия ограничиваемся графами запретов наименьшей размерности. В целом можно предположить, что чем сложнее строение $\mathcal{G}_{\min}(\eta)$, тем мощнее будет критерий (тем больше альтернатив он будет отклонять).

Кратко изложенный подход можно сформулировать так:

1. Для данного η находим $\mathbf{z}$-однородный гиперграф $\mathcal{G}_{\min}(\eta)$, где $\mathbf{z}$ — длина кратчайшего запрета частных распределений η.
2. При статистическом анализе распределения ξ оцениваем гиперграф $\mathcal{G}_{\min}(\xi)$, и если оценка изоморфна графу $\mathcal{G}_{\min}(\eta)$, то принимаем $H(\eta)$.
3. Чем меньше $\mathbf{z}$, тем меньше временная и емкостная сложность проверки критерия, а также величина $N_1(\mathbf{z}, \cdot)$.

2.2. Алгоритмы поиска некоторых номеров координат подвектора с известным либо неизвестным распределением

Везде далее будем считать, что априори справедлива гипотеза $H(\eta)$ и стоит задача определения множества Θ или его части. При известном графе $\mathcal{G}_{\min}(\eta)$ предлагается следующий последовательный алгоритм 1 поиска множества $\Theta_{\min}$.

Алгоритм 1. Поиск множества $\Theta_{\min}$

Вход: $\mathbf{x}(t) \in X^L$, $t = 1, 2, \ldots$, $\quad \mathcal{G}_{\min}(\eta)$
Выход: $\Theta_{\min}$, τ
1: $N := 1$;
2: **Пока** $\mathbf{z}_{\min}(\mathcal{X}(N)) < \mathbf{z}_{\min}(\eta)$ или $|\mathcal{G}_{\min}(\mathcal{X}(N))| > |\mathcal{G}_{\min}(\eta)|$
3: $\quad N := N + 1$;
4: **Вернуть** $\mathbf{V}(\mathcal{G}_{\min}(\mathcal{X}(N)))$ и N

Здесь в качестве статистической оценки $\Theta_{\min}$ рассматривается множество номеров координат, на которых расположены кратчайшие запреты эмпирического распределения.

Теорема 2. Пусть при гипотезе $H(\eta)$ выполнено ограничение (7). Тогда:

1. Вероятность ошибки алгоритма 1 в случае его окончания равна нулю.
2. Для распределения момента τ окончания работы алгоритма 1 справедлива оценка $\mathbf{P}\{\tau > N_1(\mathbf{z}_{\min}(\eta), \xi, \alpha)\} \leqslant \alpha$.

Доказательство.

1. Как и при доказательстве п. 2 теоремы 1, условия $H(\eta)$ и (7) обеспечивают выполнение двух первых соотношений в цепочке

$$\mathcal{G}_{\min}(\eta) \cong \mathcal{G}_{\min}(\xi_\Theta) = \mathcal{G}_{\min}(\xi) \subset \mathcal{G}_{\min}(\mathcal{X}). \tag{9}$$

Из этих условий и первого условия п. 2 алгоритма 1 также следует равенство

$$\mathbf{z}_{\min}(\xi) = \mathbf{z}_{\min}(\eta) = \mathbf{z}_{\min}(\mathcal{X}(N)),$$

которое с учётом (2) даёт последнее включение в (9).

Тогда второе условие п. 2 алгоритма означает равенство количеств рёбер левого и правого графов в цепочке (9). Отсюда следует, что $\mathcal{G}_{\min}(\xi) = \mathcal{G}_{\min}(\mathcal{X})$. Поэтому множества вершин, покрытых рёбрами каждого из них, совпадают.

2. Оценка вероятности следует из (4) и импликаций

$$\big(\mathcal{G}_{\min}(\mathcal{X}) = \mathcal{G}_{\min}(\xi)\big) \Longrightarrow \big(\mathcal{G}_{\min}(\mathcal{X}) \cong \mathcal{G}_{\min}(\eta)\big) \Longrightarrow \begin{pmatrix} \mathbf{z}_{\min}(\mathcal{X}) = \mathbf{z}_{\min}(\eta), \\ |\mathcal{G}_{\min}(\mathcal{X})| = |\mathcal{G}_{\min}(\eta)| \end{pmatrix},$$

справедливых при условиях $H(\eta)$ и (7). ■

Заметим, что если $\Theta_{\min} \neq \Theta$ (что эквивалентно условию $\mathbf{z}_{\min}(\eta) < \mathbf{z}_{\max}(\eta)$), то требуются дополнительные действия для поиска остальных элементов особенного множества.

Далее переходим к алгоритмам поиска части Θ при отсутствии информации о графе $\mathcal{G}_{\min}(\eta)$ и графах запретов большей размерности. В отличие от алгоритма 1, они работают на фиксированном объёме материала, но вероятность их ошибки в общем случае ненулевая.

Алгоритм 2 на фиксированном объёме материала N находит все кратчайшие запреты эмпирического распределения и возвращает множество вершин соответствующего графа в качестве статистической оценки части особенного множества.

Алгоритм 2. Статистическое оценивание $\Theta_{\min}$

Вход: $\mathbf{x}(t) \in X^L$, $1 \leqslant t \leqslant N$
Выход: $\Theta^*_{\min}$
1: **Вернуть** $\mathbf{V}(\mathcal{G}_{\min}(\mathcal{X}(N)))$

Следствие 1. Если при условиях $H(\eta)$ и (7) выполнено $N \geqslant N_1(\mathbf{z}_{\min}(\eta), \xi, \alpha)$, то $\mathbf{P}\{\Theta^*_{\min} = \Theta_{\min}\} \geqslant 1 - \alpha$.

Доказательство. Обозначая $\mathbf{z} = \mathbf{z}_{\min}(\eta)$, из предположения (4) с учётом ограничения (7) получаем, что событие $\{\mathcal{G}(r, \mathcal{X}) = \varnothing,\ 1 \leqslant r < \mathbf{z},\ \mathcal{G}(\mathbf{z}, \mathcal{X}) = \mathcal{G}_{\min}(\xi)\}$ происходит с вероятностью не менее $1 - \alpha$ при $N \geqslant N_1(\mathbf{z}, \xi, \alpha)$. Оно влечёт равенство $\mathcal{G}_{\min}(\mathcal{X}) = \mathcal{G}_{\min}(\xi)$, которое с учётом цепочки (9) доказательства теоремы 2 даёт равенство $\mathcal{G}_{\min}(\mathcal{X}) = \mathcal{G}_{\min}(\xi_\Theta)$. ■

Далее аналогично можем строить статистические оценки множества

$$\Theta_{\leqslant R} := \{i \in \Theta : \mathbf{z}(i, \xi_\Theta) \leqslant R\} \subset \Theta$$

номеров координат, устойчивость к запретам которых не превосходит R, — алгоритм 3.

Алгоритм 3. Построение оценки $\Theta^*_{\leqslant R}$

Вход: $\mathbf{x}(t) \in X^L$, $1 \leqslant t \leqslant N$, $R \geqslant 1$ — параметр алгоритма
Выход: $\Theta^*_{\leqslant R}$
1: **Вернуть** $\Theta^*_{\leqslant R} := \bigcup\limits_{1 \leqslant r \leqslant R} \mathbf{V}(\mathcal{G}(r, \mathcal{X}(N)))$

При некотором ограничении, в общем случае усиливающем ограничение (7), докажем следующую оценку надёжности алгоритма 3.

Следствие 2. Если $N \geqslant N_1(R,\xi,\alpha)$, выполнено условие $H(\eta)$ и ограничение $\forall i \notin \Theta\ (\mathbf{z}(i,\xi) > R)$, то $\mathbf{P}\left\{\Theta^*_{\leqslant R} = \Theta_{\leqslant R}\right\} \geqslant 1-\alpha$.

Доказательство. Согласно ограничению, для каждого $1 \leqslant r \leqslant R$ рёбра графа $\mathcal{G}(r,\xi)$ не содержат номеров из $\bar{\Theta}$, откуда $\Theta_{\leqslant R} = \bigcup\limits_{1\leqslant r\leqslant R} \mathbf{V}(\mathcal{G}(r,\xi))$.

Остаётся заметить, что при $N \geqslant N_1(R,\xi,\alpha)$, согласно предположению (4), событие $\{\mathcal{G}(r,\mathcal{X}) = \mathcal{G}(r,\xi),\ 1 \leqslant r \leqslant R\}$ происходит с вероятностью не менее $1-\alpha$. С учётом предыдущего равенства оно влечёт событие $\left\{\Theta^*_{\leqslant R} = \Theta_{\leqslant R}\right\}$. ■

Не будем останавливаться на оценках сложности методов построения эмпирических простых запретов, а также на оценках сложности проверки условия изоморфизма r-графов и особенностях реализации этих методов — этот круг проблем выходит за рамки работы, причём проблема изоморфизма хорошо известна. Сосредоточимся на вопросах построения графов запретов и оценках числа наблюдений.

3. Расчет параметров критерия и алгоритмов

Для применения критерия (8) согласия с $H(\eta)$ и алгоритмов 1–3 поиска части Θ надо уметь:

1) рассчитывать значение $\mathbf{z}_{\min}(\eta)$, строить граф $\mathcal{G}_{\min}(\eta)$;
2) строить функции $N_1(r,\xi,\alpha)$ и проверять ограничение (7).

Согласно теореме 1, знание функций $N_1(\cdot)$ требуется для оценки надёжности критерия при его применении. Такое же замечание справедливо для алгоритмов 2 и 3, согласно следствиям из теоремы 2.

3.1. Формула для N_1

Получим формулу для оценочной функции $N_1(r,\xi,\alpha)$ в случае независимых наблюдений $\mathbf{x}(t) \sim \xi$. Обозначим далее через

$$p_{\min}(r,\xi) = \min\left\{p = \mathbf{P}\{\xi_{\mathbf{J}} = \mathbf{a}\} : J \in \{1,\ldots,L\}^{(r)},\ \mathbf{a} \in X^r,\ p > 0\right\} > 0$$

самую малую ненулевую вероятность r-мерных распределений ξ.

Теорема 3. Пусть $\mathbf{x}(t) \sim \xi$, $t \geqslant 1$ — последовательность независимых случайных векторов, $0 < \alpha < 1$. Тогда в качестве функции $N_1(\cdot)$ может быть взята функция

$$N_1^*(r,\xi,\alpha) = \frac{1}{p_{\min}(r,\xi)}\left(r\ln\frac{kLe}{r} - \ln\alpha\right).$$

Доказательство. Достаточно доказать, что при $N \geqslant N_1^*(\cdot)$ справедлива оценка (3).

Из неравенства $r! > \left(\frac{r}{e}\right)^r$ имеем $\binom{L}{r} < \frac{L^r}{r!} < \left(\frac{Le}{r}\right)^r$. Положим $p = p_{\min}(r,\xi)$. Используя эту оценку и неравенство $1+x \leqslant e^x$, оцениваем сверху вероятность непоявления какого-либо допустимого подвектора на каких-либо r координатах величиной

$$\mathbf{P}\left\{\exists J \in \{1,\ldots,L\}^{(r)}\ (\mathcal{X}(N)_J \neq \mathrm{Supp}(\xi_J))\right\} \leqslant$$

$$\leqslant \sum_{J\in\{1,\ldots,L\}^{(r)}} \sum_{\mathbf{a}\in\xi_J(\Omega)} \mathbf{P}\left\{\mathbf{x}_J(t) \neq \mathbf{a},\ 1 \leqslant t \leqslant N\right\} \leqslant k^r\binom{L}{r}(1-p)^N < \left(\frac{kLe}{r}\right)^r \exp(-Np),$$

которая не превосходит α при $N \geqslant N_1^*(r,\xi,\alpha)$. ■

Поясним наличие величины $1/p$, $p = p_{\min}(r, \xi)$, в предложенном выражении для N_1. Пусть ν — случайная величина, равная числу наблюдений до первого момента появления исхода с вероятностью p. В схеме независимых наблюдений она имеет геометрическое распределение с параметром p. Для момента τ первого появления всех возможных r-грамм в $\mathcal{X}$ справедлива оценка $\tau \geqslant \nu$. Поэтому для момента остановки τ последовательного алгоритма, ожидающего появления всех возможных r-грамм в $\mathcal{X}$, имеем $\tau \geqslant \nu$. Отсюда получаем $\mathbf{E}\tau \geqslant \mathbf{E}\nu = 1/p$.

Так как

$$\mathbf{P}\{\tau > N\} \geqslant \mathbf{P}\{\nu > N\} = p(1-p)^N + p(1-p)^{N+1} + \ldots = (1-p)^N,$$

вероятность $\mathbf{P}\{\tau > N\}$ может быть не больше α только при $N \geqslant \dfrac{\ln \alpha}{\ln(1-p)}$. Последняя дробь эквивалентна $\dfrac{1}{p} \ln \dfrac{1}{\alpha}$ при $p \to 0$. Поэтому порядок величины $1/p$ в предложенной функции $N_1^*(\cdot)$ не может быть уменьшен.

Теперь можем получить оценку объёма материала, при котором критерий идентификации (8) имеет вероятность ошибки не более заданной величины при гипотезе о равномерности частных распределений ξ. Она очевидно вытекает из теоремы 3 и п. 2 теоремы 1.

Следствие 3. Пусть в схеме независимых наблюдений выполнено

$$\mathbf{z} = \mathbf{z}_{\min}(\eta) < \infty, \quad N \geqslant N_1^*(\mathbf{z}, \xi, \alpha) = 2^{\mathbf{z}} \left(\mathbf{z} \ln \frac{kLe}{\mathbf{z}} - \ln \alpha \right).$$

Тогда при альтернативе о равномерности всех $\mathbf{z}$-мерных частных распределений ξ (сюда, в частности, входит простая гипотеза $\xi \sim U(X^L)$) вероятность ошибки критерия (8) не превосходит $1 - \alpha$.

Получим оценки для $p_{\min}(r, \xi)$, считая по определению $p_{\min}(0, \xi) = 1$.

Теорема 4. Для любого $1 \leqslant r \leqslant \mathbf{z} - 1$, $\mathbf{z} = \mathbf{z}_{\min}(\xi)$, справедливы неравенства

$$p_{\min}(r, \xi) \leqslant \frac{1}{k}\, p_{\min}(r-1, \xi) \leqslant \ldots \leqslant \frac{1}{k^r}\, p_{\min}(0, \xi) = k^{-r}.$$

Следовательно, $p_{\min}(\mathbf{z}, \xi) \leqslant p_{\min}(\mathbf{z} - 1, \xi) \leqslant k^{1-\mathbf{z}}$.

Доказательство. Заметим, что при $r < \mathbf{z}$ каждая вероятность $(r-1)$-мерного распределения ξ равна сумме некоторых k ненулевых вероятностей r-мерных распределений и поэтому не меньше величины $k p_{\min}(r, \xi)$. Отсюда вытекает первое неравенство в первой цепочке, а из него остальные. Во второй цепочке первое неравенство очевидно, а второе получено из первой цепочки при $r = \mathbf{z} - 1$. ■

Таким образом, величина $p_{\min}(r, \xi)$ при $0 \leqslant r \leqslant \mathbf{z} - 1$ убывает от начального значения 1 при каждом увеличении r на 1 не менее чем в k раз; равенство $p_{\min}(r, \xi) = k^{-r}$ достигается тогда и только тогда, когда все r-мерные распределения ξ равномерны. Величина $p_{\min}(\mathbf{z}, \xi)$ может быть как меньше, так и больше величины $k^{-1} p_{\min}(\mathbf{z} - 1, \xi)$.

3.2. Упрощающие условия

Покажем, что условие

$$\text{случайные векторы } \xi_\Theta \text{ и } \xi_{\bar{\Theta}} \text{ независимы} \tag{10}$$

существенно упрощает вычисление $\mathbf{z}_{\min}(\xi)$ и проверку ограничения (7).

Теорема 5. Пусть для некоторого $\varnothing \neq A \subset \{1, \ldots, L\}$ случайные векторы ξ_A и $\xi_{\bar{A}}$ независимы. Тогда

$$\begin{aligned} &1)\ \mathcal{G}(r, \xi) = \mathcal{G}(r, \xi_A) \sqcup \mathcal{G}(r, \xi_{\bar{A}}) \ \text{ для всех } \ 1 \leqslant r \leqslant L; \\ &2)\ \mathbf{z}(a, \xi) = \mathbf{z}(a, \xi_A) \ \text{ для всех } \ a \in A; \\ &3)\ \text{если } \ A = \{a\}, \ \text{ то } \ \mathbf{z}(a, \xi) = \begin{cases} \infty \text{ при } \operatorname{Supp}(\xi_a) = X, \\ 1 \text{ иначе.} \end{cases} \end{aligned} \tag{11}$$

Доказательство.

1. Предположим противное: пусть существует $\mathbf{b}$ — простой запрет распределения $\xi_{I \cup J}$, где $\varnothing \neq I \subset A$; $\varnothing \neq J \subset \bar{A}$. Тогда из условия независимости имеем

$$0 = \mathbf{P}\{\xi_{I \cup J} = \mathbf{b}\} = \mathbf{P}\{\xi_I = \mathbf{b}_I\}\mathbf{P}\{\xi_J = \mathbf{b}_J\},$$

и один из сомножителей в последнем выражении равен нулю. Но это противоречит простоте запрета. Пункт 1 доказан.

2. Пункт 2 следует из п. 1, поскольку вершина $a \in A$ может лежать только в рёбрах графов $\mathcal{G}(r, \xi_A)$.

3. Если $\operatorname{Supp}(\xi_a) \neq X$, то утверждение очевидно. В противном случае распределение ξ_a не имеет запретов, а в r-мерных простых запретах при $r \geqslant 2$ номер a не участвует согласно п. 1. Пункт 3 доказан. ■

Равенство (11) означает, что если ξ можно разделить на два независимых подвектора, то графы запретов распадаются на не связанные между собою графы запретов подвекторов. Можно также показать, что условие независимости в теореме 5 не является необходимым для условия (11).

Следствие 4. Если при справедливости гипотезы $H(\eta)$ выполнено условие (10), то ограничение (7) равносильно условию $\mathbf{z}_{\min}(\eta) < \mathbf{z}_{\min}(\xi_{\bar{\Theta}})$.

Доказательство. Из п. 2 теоремы 5 имеем $\min\limits_{i \in \Theta} \mathbf{z}(i, \xi) = \mathbf{z}_{\min}(\xi_\Theta)$, $\min\limits_{i \in \bar{\Theta}} \mathbf{z}(i, \xi) = \mathbf{z}_{\min}(\xi_{\bar{\Theta}})$. Осталось заметить, что при гипотезе $H(\eta)$ выполнено равенство $\mathbf{z}_{\min}(\xi_\Theta) = \mathbf{z}_{\min}(\eta)$. ■

Введём второе упрощающее условие

$$\xi_\Theta \ \text{ не зависит от } \ \xi_{\bar{\Theta}} \sim U(X^{L - |\Theta|}). \tag{12}$$

Оно получено путём добавления в условие независимости (10) условия равномерности распределения координат, не принадлежащих подвектору. Можно сказать, что в этом случае подвектор на координатах с номерами из Θ «погружен» в не зависящий от него равномерно распределённый случайный вектор $\xi_{\bar{\Theta}}$.

При условии (12) из п. 3 теоремы 5 получаем, что $\mathbf{z}_{\min}(\xi_{\bar{\Theta}}) = \infty$. Тогда, согласно следствию 4, выполнено ограничение (7) и корректно далее используемое единое обозначение для устойчивостей к запретам трёх распределений

$$\mathbf{z} = \mathbf{z}_{\min}(\xi) = \mathbf{z}_{\min}(\xi_\Theta) = \mathbf{z}_{\min}(\eta).$$

Кроме того, при условии (12) для любого R выполнено ограничение следствия 2, при котором оценивалась надёжность алгоритма 3.

При введённом условии (12) и гипотезе $H(\eta)$ распределение ξ определяется распределением η (с точностью до перестановки координат). Поэтому можем явно выразить фигурирующую в оценках объёма материала величину $p_{\min}(\mathbf{z}, \xi)$ через распределение η.

Теорема 6. Если выполнены условия $H(\eta)$, (12) и $\mathbf{z} < L$, то

$$p_{\min}(r,\xi) = \min\{p_{\min}(r,\eta), \frac{1}{k}\, p_{\min}(r-1,\eta)\},\ 1 \leqslant r \leqslant \mathbf{z}; \tag{13}$$

$$p_{\min}(r,\xi) \leqslant k^{-r},\ 1 \leqslant r \leqslant \mathbf{z}. \tag{14}$$

Доказательство. Докажем утверждения в случае $r = 1$. Здесь имеем

$$p_{\min}(1,\xi) = \min\{p_{\min}(1,\xi_\Theta), p_{\min}(1,\xi_{\bar\Theta})\},$$

что совпадает с правой частью (13), поскольку $p_{\min}(1,\xi_{\bar\Theta}) = \dfrac{1}{k} = \dfrac{1}{k}\, p_{\min}(0,\eta)$. Из последней формулы также вытекает справедливость (14) при $r = 1$.

Осталось доказать утверждения в случае $2 \leqslant r \leqslant L-1$, $\mathbf{z} \geqslant 2$. Очевидно, что для всех $2 \leqslant r \leqslant L-1$ выполнено равенство

$$p_{\min}(r,\xi) = \min\{p_{\min}(r,\xi_\Theta), p_{\min}(r,\xi_{\bar\Theta}), p_{\mathrm{joi}}(r)\}, \tag{15}$$

где $p_{\mathrm{joi}}(r)$ — минимум ненулевых вероятностей r-мерных распределений на координатах, содержащих номера из Θ и $\bar\Theta$ одновременно.

Используя равномерность распределения $\xi_{\bar\Theta}$ в первом переходе и первую цепочку неравенств теоремы 4 во втором переходе, при $2 \leqslant r \leqslant \mathbf{z}$ имеем равенства

$$p_{\mathrm{joi}}(r) = \min_{1\leqslant l\leqslant r-1} p_{\min}(l,\xi_\Theta)k^{l-r} = p_{\min}(r-1,\xi_\Theta)k^{-1} = p_{\min}(r-1,\eta)k^{-1}.$$

Согласно (15) и полученному выражению для $p_{\mathrm{joi}}(\mathbf{z})$, для доказательства равенства (13) осталось обосновать неравенство

$$p_{\min}(r-1,\eta)k^{-1} \leqslant p_{\min}(r,\xi_{\bar\Theta}) = k^{-r}.$$

Но при всех $2 \leqslant r \leqslant \mathbf{z}$ оно следует из первой цепочки теоремы 4.

Из последнего неравенства также вытекает (14). ■

Заметим, что при $\mathbf{z} = L$ утверждения теоремы могут быть неверны. В частности, тогда в (13) левая часть равна первому выражению под знаком минимума, а второе выражение под знаком минимума может быть меньше первого. Например, для $\xi = \eta \equiv (0)$ имеем $p_{\min}(1,\xi) = 1 > \dfrac{1}{k} = \dfrac{1}{k}\, p_{\min}(0,\eta)$.

О возможности обобщения результатов на случай $r > \mathbf{z}$ заметим следующее. Равенство (13) прямого обобщения не допускает, поскольку, например, при $\mathbf{z} = M$ величина $p_{\min}(r,\eta)$ не определена для $r > \mathbf{z}$. Неравенство (14) при $r > \mathbf{z}$ в общем случае неверно. Например, при $\xi = (0,x)$, $x \sim U(X)$ имеем $\mathbf{z} = 1$, $p_{\min}(2,\xi) = \dfrac{1}{k} > k^{-2}$.

Итак, при гипотезе $H(\eta)$ дополнительное условие (12) позволяет по величинам, полностью определяемым распределением η:

— построить критерий и алгоритм 1;

— рассчитать оценку $N_1^*(\mathbf{z}_{\min}(\eta),\xi,\alpha)$ объёма материала критерия и алгоритмов 1 и 2.

Заметим, что при этом условии из (14) вытекает нижняя оценка для достаточного числа наблюдений: для всех $1 \leqslant r \leqslant \mathbf{z}_{\min}(\eta)$

$$N_1^*(r,\xi,\alpha) \geqslant k^r \left(r\left(\ln\frac{kL}{r} + 1\right) - \ln\alpha \right).$$

Заключение

Определение запрета размерности r на множестве номеров координат $J \in \in \{1,\dots,L\}^{(r)}$, приведённое в работе, обобщает определение 1 [2, с. 55], в котором такие запреты рассматриваются лишь при $J = \{1,\dots,r\}$. Поэтому наше понятие кратчайшего запрета отличается от понятий наименьшего и минимального запретов [2, с. 55]. В наших терминах $\mathbf{a} \in X^s$ является минимальным запретом в смысле [2], если он является запретом на множестве номеров координат $J = \{n-s+1,\dots,n\}$ (последних s координатах) и не является запретом на множестве $\{n-s+2,\dots,n\}$.

В [2, с. 55] предложен алгоритм построения статистической оценки множества всех минимальных запретов распределения ξ размерности не более s_0 при независимых наблюдениях. Получено выражение для математического ожидания числа эмпирических запретов (на последних координатах) размерности не более s_0, а также верхняя оценка для математического ожидания числа $\nu(s_0)$ таких эмпирических запретов, не являющихся теоретическими:

$$\mathbf{E}\nu(s_0) \leqslant s_0 k^{s_0}(1 - p'_{\min}(s_0,\xi))^N,$$

где $p'_{\min}(s_0,\xi)$ — минимум ненулевых вероятностей распределения на последних s_0 координатах. Очевидно, что эта величина связана с аналогичной введённой величиной неравенством $p'_{\min}(s_0,\xi) \geqslant p_{\min}(s_0,\xi)$. Из неравенства для $\mathbf{E}\nu(s_0)$ с помощью неравенства Маркова сделан вывод о состоятельности при $N \to \infty$ предложенной оценки множества таких запретов.

Ранее в [5] и других работах А. А. Грушо и Е. Е. Тимониной исследовалась задача построения состоятельных критериев, которые предполагалось применять для статистического выявления сбоев в протоколах и технических устройствах.

Проведённая конкретизация задачи и развитие понятийного аппарата позволили сформулировать конкретные алгоритмы её решения, дать допредельные оценки вероятностей ошибок. Такое направление развития предполагалось в заключении работы [1].

Предложенный подход, в свою очередь, может развиваться в следующих направлениях:

— получение явных верхних оценок числа наблюдений для критерия и алгоритмов в случае «функциональной вставки» $\eta \sim \mathbf{xf}$, в том числе для конкретных классов функций;
— получение оценок числа наблюдений для более сложных распределений выборки, например схемы конечно зависимых наблюдений, схемы псевдослучайного образования аргументов функции.

ЛИТЕРАТУРА

1. *Грушо А. А., Тимонина Е. Е.* Запреты в дискретных вероятностно-статистических задачах // Дискретная математика. 2011. Т. 23. № 2. С. 53–58.
2. *Грушо А. А., Грушо Н. А., Тимонина Е. Е.* Статистические методы определения запретов вероятностных мер на дискретных пространствах // Информатика и её применение. 2013. Т. 7. № 1. С. 54–57.
3. *Михайлов В. Г., Чистяков В. П.* О задачах теории конечных автоматов, связанных с числом прообразов выходной последовательности // Обозрение прикл. и промышл. матем. Сер. дискретн. матем. 1994. Т. 1. Вып. 1. С. 7–32.

4. *Сумароков С. Н.* Запреты двоичных функций и обратимость для одного класса кодирующих устройств // Обозрение прикл. и промышл. матем. Сер. дискретн. матем. 1994. Т. 1. Вып. 1. С. 33–55.
5. *Грушо А. А., Тимонина Е. Е.* Некоторые связи между дискретными статистическими задачами и свойствами вероятностных мер на топологических пространствах // Дискретная математика. 2006. Т. 18. № 4. С. 128–136.

REFERENCES

1. *Grusho A. A., Timonina E. E.* Zaprety v diskretnykh veroyatnostno-statisticheskikh zadachakh [Prohibitions in discrete probabilistic statistical problems.] Diskretnaya Matematika, 2011, vol. 23, no. 2, pp. 53–58. (in Russian)
2. *Grusho A. A., Grusho N. A., Timonina E. E.* Statisticheskie metody opredeleniya zapretov veroyatnostnykh mer na diskretnykh prostranstvakh [Statistical techniques of bans determination of probability measures in discrete spaces.] Inform. Primen., 2013, vol. 7, no. 1, pp. 54–57. (in Russian)
3. *Mikhaylov V. G., Chistyakov V. P.* O zadachakh teorii konechnykh avtomatov, svyazannykh s chislom proobrazov vykhodnoy posledovatel'nosti [Problems of the finite automata theory associated with a number of inverse images of the output sequence.] Obozrenie Prikl. i Promyshl. Matem. Ser. Diskretn. Matem., 1994, vol. 1, iss. 1, pp. 7–32. (in Russian)
4. *Sumarokov S. N.* Zaprety dvoichnykh funktsiy i obratimost' dlya odnogo klassa kodiruyushchikh ustroystv [Prohibitions of binary functions and reversibility for a class of encoders.] Obozrenie Prikl. i Promyshl. Matem. Ser. Diskretn. Matem., 1994, vol. 1, iss. 1, pp. 33–55. (in Russian)
5. *Grusho A. A., Timonina E. E.* Nekotorye svyazi mezhdu diskretnymi statisticheskimi zadachami i svoystvami veroyatnostnykh mer na topologicheskikh prostranstvakh [Some relations between discrete statistical problems and properties of probability measures on topological spaces.] Diskretnaya Matematika, 2006, vol. 18, no. 4, pp. 128–136. (in Russian)

УДК 519.214

АППРОКСИМАЦИЯ РАСПРЕДЕЛЕНИЯ ЧИСЛА МОНОТОННЫХ ЦЕПОЧЕК ЗАДАННОЙ ДЛИНЫ В СЛУЧАЙНОЙ ПОСЛЕДОВАТЕЛЬНОСТИ СЛОЖНЫМ РАСПРЕДЕЛЕНИЕМ ПУАССОНА

А. А. Минаков

Московский государственный технический университет радиотехники, электроники и автоматики (МИРЭА), г. Москва, Россия

Рассматривается распределение числа монотонных цепочек заданной длины s в последовательности из n независимых равномерно распределённых на множестве $\{0,\dots,N-1\}$ случайных величин с фиксированным числом исходов N. С помощью метода Стейна получена оценка расстояния по вариации между распределением числа монотонных цепочек длины s и сложным пуассоновским распределением. На основании оценки доказана предельная теорема для числа монотонных цепочек при $n, s \to \infty$. В теореме аппроксимирующим распределением является распределение суммы пуассоновского числа независимых случайных величин, имеющих геометрическое распределение.

Ключевые слова: *монотонные цепочки, оценка точности сложной пуассоновской аппроксимации, сложное пуассоновское распределение, метод Стейна.*

DOI 10.17223/20710410/28/2

COMPOUND POISSON APPROXIMATION OF THE NUMBER DISTRIBUTION FOR MONOTONE STRINGS OF FIXED LENGTH IN A RANDOM SEQUENCE

A. A. Minakov

Moscow State Institute of Radio Engineering, Electronics and Automation, Moscow, Russia

E-mail: minak-ski@yandex.ru

We study the number distribution for monotone strings of a length s in a sequence of n random independent variables uniformly distributed on the set $\{0,\dots,N-1\}$ where N is a constant. By means of the Stein method we construct an estimate of the variation distance between this distribution and a compound Poisson distribution. As a corollary of this result we prove the limit theorem as $n, s \to \infty$ for the number of monotone strings. The approximating distribution is the distribution of the sum of Poisson number of independent random variables with geometric distribution.

Keywords: *monotone strings, estimate of the variation distance of the compound Poisson approximation, compound Poisson distribution, Stein method.*

Введение

Пусть $X_1, X_2, \dots, X_n$ — отрезок последовательности, состоящей из независимых случайных величин, каждая из которых имеет равномерное распределение на множестве $\{0,\dots,N-1\}$.

Определение 1. Монотонной цепочкой длины s, $s \in \mathbb{N}$, с началом в t назовём событие $E_t = \{X_t \leqslant X_{t+1} \leqslant \ldots \leqslant X_{t+s-1}\}$.

Определение 2. Монотонной серией длины s, $s \in \mathbb{N}$, с началом в t назовём событие $Y_t = \{X_{t-1} > X_t \leqslant X_{t+1} \leqslant \ldots \leqslant X_{t+s-1} > X_{t+s}\}$.

Введём случайную величину

$$\xi_n(s) = \sum_{t=1}^{n} I\{E_t\},$$

равную числу монотонных цепочек длины s, которые начинаются на отрезке $X_1, X_2, \ldots, X_n$. Для избежания краевого эффекта и облегчения вычислений предполагаем, что рассматривается бесконечная в обе стороны последовательность $\{X_a : a \in \mathbb{Z}\}$. Через $I\{A\}$ обозначаем индикатор события A.

В. Л. Гончаров [1] доказал несколько предельных теорем для монотонных серий в двоичной последовательности, рассмотрев чередование событий в ряде независимых опытов, отвечающих схеме Бернулли. J. Wolfowitz [2] доказал условия сходимости распределения числа монотонных серий заданной длины в конечной бесповторной последовательности к распределению Пуассона и стандартному нормальному распределению. F. N. David и D. E. Barton [3] доказали условия для пуассоновской аппроксимации числа монотонных серий длины больше заданной в конечной бесповторной последовательности. Их результаты обобщил B. G. Pittel [4], который доказал теорему о сходимости распределения числа монотонных серий длины больше заданной к распределению Пуассона. O. Chryssaphinou, S. Papastavridis и E. Vaggelatou [5] доказали теорему об аппроксимации распределения числа монотонных серий заданной длины в стационарной цепи Маркова пуассоновским распределением. Н. М. Меженная [6] доказала многомерную нормальную теорему для числа монотонных серий заданной длины. В данной работе находится оценка расстояния по вариации между распределением числа монотонных цепочек длины s и сложным пуассоновским распределением.

1. Оценка по вариации и предельная теорема

Введём некоторые обозначения. Условимся обозначать $d(\Phi, \Psi)$ расстояние по вариации между распределениями Φ и Ψ. Для распределений Φ и Ψ на множестве $\{0, 1, \ldots\}$ справедлива следующая формула (теорема Шеффе):

$$d(\Phi, \Psi) = \frac{1}{2} \sum_{m=0}^{\infty} |\Psi\{m\} - \Phi\{m\}|.$$

Распределение случайной величины ζ будем обозначать $L(\zeta)$.

Пусть $\Lambda = (\lambda_1, \lambda_2, \ldots)$ — последовательность неотрицательных действительных чисел, причём сходится ряд $\sum\limits_{k=1}^{\infty} \lambda_k < \infty$. Пусть $\{\theta_1, \theta_2, \ldots\}$ — последовательность независимых случайных величин, причём случайная величина θ_k имеет распределение Пуассона с параметром λ_k, где $k \in \mathbb{N}$. Распределение случайной величины $\sum\limits_{k=1}^{\infty} k\theta_k$ называется сложным распределением Пуассона, которое будем обозначать $CP(\Lambda)$.

Введём несколько определений аналогично работе [6]. Пусть A_s — число неубывающих цепочек длины s из символов алфавита $\{0, \ldots, N-1\}$. Тогда для $s, N \in \mathbb{N}$ справедлива формула

$$A_s = \binom{s+N-1}{s}. \tag{1}$$

Пусть B_s — число цепочек длины $s+1$, не являющихся неубывающими, но становящихся таковыми после удаления первого элемента. Тогда для $s, N \in \mathbb{N}$ справедлива формула

$$B_s = \left(\begin{array}{c} s+N \\ s+1 \end{array} \right) \frac{s\,(N-1)}{N+s}.$$

Пусть C_s — число цепочек длины $s+2$, не являющихся неубывающими, но становящихся таковыми после удаления первого и последнего элементов. Тогда для $N \geqslant 3$ и $s \in \mathbb{N}$ справедлива формула

$$C_s = \left(\begin{array}{c} s+N-1 \\ s+2 \end{array} \right) \frac{N\,(s^2+s-1) - s^2 - s}{N-2}.$$

Далее положим

$$\lambda_\nu = \begin{cases} \dfrac{nC_{s+\nu-1}}{N^{s+\nu+1}} & \text{для } \nu \in \{1, \ldots, s-1\}, \\ \dfrac{n\,(2B_{s+\nu-1}N + (2s-2-\nu)\,C_{s+\nu-1})}{\nu N^{s+\nu+1}} & \text{для } \nu \in \{s, \ldots, 2s-2\}, \\ \dfrac{nA_{2s-1}}{(2s-1)\,N^{3s-2}} & \text{для } \nu \in \{2s-1\}, \\ 0 & \text{для } \nu \in \{2s, \infty\}. \end{cases} \tag{2}$$

На основе метода Стейна и результатов работы [7] докажем следующую теорему.

Теорема 1. Пусть $(X_1, X_2, \ldots, X_n)$ — отрезок последовательности, состоящей из независимых случайных величин, каждая из которых имеет равномерное распределение на множестве $\{0, \ldots, N-1\}$, $N = \mathrm{const} \geqslant 3$ и все λ_ν имеют вид (2). Тогда

$$\begin{aligned} &d\left(L\left(\xi_n\left(s\right)\right), CP\left(\lambda_1, \lambda_2, \ldots, \lambda_{2s-1}, 0, 0, \ldots\right)\right) \leqslant \\ &\leqslant \exp\left\{\sum_{k=1}^{2s-1} \lambda_k\right\} \frac{n\,(6s-5)}{\left(sN^{-1}+1\right)^2 N^{2s}} \left(\begin{array}{c} s+N \\ s \end{array} \right)^2. \end{aligned} \tag{3}$$

Из теоремы 1 выведем предельную теорему для случайной величины $\xi_n\left(s\right)$.

Теорема 2. Пусть $(X_1, X_2, \ldots, X_n)$ — отрезок последовательности, состоящей из независимых случайных величин, каждая из которых имеет равномерное распределение на множестве $\{0, \ldots, N-1\}$, и $N = \mathrm{const} \geqslant 3$. Если $n, s \to \infty$ так, что

1) $s/n \to 0$,
2)
$$n\,(s+N)^{N-1}\,N^{-s-1}\,((N-2)!)^{-1} \to \lambda \in (0, \infty), \tag{4}$$

то $L\left(\xi_n\left(s\right)\right) \to CP\left(\lambda\left(1-N^{-1}\right), \lambda N^{-1}\left(1-N^{-1}\right), \lambda N^{-2}\left(1-N^{-1}\right), \ldots\right)$.

Так как N фиксировано, а $s \to \infty$, число монотонных цепочек длины s, не содержащих все символы из множества $\{0, \ldots, N-1\}$, стремится к нулю. В пределе количества монотонных цепочек длины s в монотонных сериях независимы и имеют геометрическое распределение (с параметром $1/N$). Число таких серий распределено по закону Пуассона (с параметром λ).

Предельным распределением в теореме 2 является распределение суммы пуассоновского (с параметром λ) числа независимых случайных величин, имеющих геометрическое распределение (с параметром $1/N$).

2. Доказательство теорем

Для доказательства теоремы 1 понадобится следующая теорема о суммах случайных индикаторов [7].

Пусть Γ — произвольный конечный набор индексов; I_a $(a \in \Gamma)$ — случайные индикаторы; $W = \sum\limits_{a\in\Gamma} I_a$. Для каждого I_a разделим некоторым образом множество Γ на четыре непересекающихся множества $\{a\}$, Γ_a^{vs}, Γ_a^b, Γ_a^{vw} и положим

$$U_a = \sum_{l\in\Gamma_a^{vs}} I_l, \qquad V_a = \sum_{l\in\Gamma_a^b} I_l.$$

Определим набор $\Lambda = (\lambda_1, \ldots, \lambda_{D+1}, 0, \ldots)$, где

$$\lambda_i = i^{-1} \sum_{a\in\Gamma} \mathbf{E}\left\{I_a I\left\{I_a + U_a = i\right\}\right\}, \tag{5}$$

и величину $D = \max\limits_a |\Gamma_a^{vs}|$. Введём обозначение $\varphi = \sum\limits_{a\in\Gamma} \sum\limits_{i=1}^{|\Gamma_a^{vs}|+1} \varphi_{ai}$,

где $$\varphi_{ai} = \mathbf{E}\left|\mathbf{E}\left\{I_a I\left\{I_a + U_a = i\right\}\left|(I_b : b \in \Gamma_a^{vw})\right.\right\} - \mathbf{E}\left\{I_a I\left\{I_a + U_a = i\right\}\right\}\right|. \tag{6}$$

Теорема 3. При любом выборе непересекающихся множеств $\{a\}$, Γ_a^{vs}, Γ_a^b, Γ_a^{vw} справедлива оценка

$$d\left(L\left(W\right), CP\left(\Lambda\right)\right) \leqslant c_1\left(\Lambda\right)\phi + c_2\left(\Lambda\right) \sum_{a\in\Gamma} \left(\left(\mathbf{E}I_a\right)^2 + \mathbf{E}I_a \mathbf{E}\left(U_a + V_a\right) + \mathbf{E}I_a V_a\right),$$

где $\max\left\{c_1\left(\Lambda\right),\ c_2\left(\Lambda\right)\right\} \leqslant \exp\left\{\sum\limits_{k=1}^{\infty} \lambda_k\right\}$.

Воспользуемся результатами теоремы 3. В нашем случае $\Gamma = \{1, 2, \ldots, n\}$, $I_a = I\left\{E_t\right\} = I\left\{X_t \leqslant X_{t+1} \leqslant \ldots \leqslant X_{t+s-1}\right\}$. Выберем множества Γ_a^{vs}, Γ_a^b, Γ_a^{vw} следующим образом:

$$\Gamma_t^{vs} = \left\{k \in \Gamma\backslash\{t\} : |t-k| < s\right\}, \quad \Gamma_t^{vw} = \left\{k \in \Gamma : |t-k| > 2s-2\right\},$$
$$\Gamma_t^b = \Gamma\backslash\left(\{t\} \cup \Gamma_t^{vs} \cup \Gamma_t^{vw}\right) = \left\{k \in \Gamma : s \leqslant |t-k| \leqslant 2s-2\right\}.$$

В силу этих определений $D = \max\limits_a |\Gamma_a^{vs}| = 2s-2$ и, следовательно, $D+1 = 2s-1$. В обозначениях теоремы 3 для $t \in \{1, \ldots, n\}$ имеем

$$U_t = \sum_{k=t-s+1, k\neq t}^{t+s-1} I\left\{E_k\right\}; \tag{7}$$

$$V_t = \sum_{k=t-2s+2}^{t-s} I\left\{E_k\right\} + \sum_{k=t+s}^{t+2s-2} I\left\{E_k\right\}. \tag{8}$$

Согласно равенству (6), для $I_a = I\left\{E_t\right\}$ получаем

$$\varphi_{ai} = \mathbf{E}\left|\mathbf{E}\left\{I_a I\left\{I_a + U_a = i\right\}\left|(I_b : b \in \Gamma_a^{vw})\right.\right\} - \mathbf{E}\left\{I_a I\left\{I_a + U_a = i\right\}\right\}\right| = 0,$$

и из определения ϕ вытекает

$$\phi = \sum_{a\in\Gamma} \sum_{i=1}^{|\Gamma_a^{vs}|+1} \phi_{ai} = 0. \tag{9}$$

По формулам (1) и (7) для $t \in \{1, \ldots, n\}$ получаем

$$\mathbf{E}U_t = \sum_{k=t-s+1, k\neq t}^{t+s-1} \mathbf{P}\{E_k\} = (2s-2)\, A_s N^{-s} = (2s-2) \left(\begin{array}{c} s+N-1 \\ s \end{array} \right) N^{-s}. \tag{10}$$

По формулам (1) и (8) для $t \in \{1, \ldots, n\}$ нетрудно получить

$$\begin{aligned} \mathbf{E}V_t = \sum_{k=t-2s+2}^{t-s} \mathbf{P}\{E_k\} + \sum_{k=t+s}^{t+2s-2} \mathbf{P}\{E_k\} = (2s-2)\, A_s N^{-s} = \\ = (2s-2) \left(\begin{array}{c} s+N-1 \\ s \end{array} \right) N^{-s}. \end{aligned} \tag{11}$$

С помощью равенств (10) и (11) получаем

$$\mathbf{E}\left(U_t + V_t\right) = (4s-4)\, A_s N^{-s}. \tag{12}$$

В соответствии со способом разбиения множества Γ случайные величины $I\{E_t\}$ и V_t независимы. Следовательно,

$$\mathbf{E}\left(I\{E_t\}\, V_t\right) = \mathbf{P}\{E_t\} \cdot \mathbf{E}V_t. \tag{13}$$

По формуле (5) для любого $\nu \in \{1, \ldots, 2s-1\}$ имеем

$$\lambda_\nu = \nu^{-1} \sum_{t=1}^{n} \mathbf{E}\left(I\{E_t\}\, I\{I\{E_t\} + U_t = \nu\}\right). \tag{14}$$

Обозначим $\alpha_t(\nu) \equiv I\{E_t\}\, I\{I\{E_t\} + U_t = \nu\}$. При фиксированном t, согласно (7), случайная величина $\alpha_t(\nu)$ равна 1 лишь в том случае, когда в отрезке $(X_{t-s+1}, \ldots, X_{t+2s-2})$ длины $3s-2$ встретились монотонная цепочка длины s с началом в t и ещё ровно $\nu - 1$ монотонных цепочек длины s. Наличие монотонной цепочки длины s с началом в t не позволяет без перекрытия с ней расположиться другой монотонной цепочке длины s. Следовательно, все ν цепочек образуют на отрезке $(X_{t-s+1}, \ldots, X_{t+2s-2})$ одну монотонную цепочку длины $s+\nu-1$.

Для вычисления выражения (14) при $\nu \in \{1, \ldots, s-1\}$ подсчитаем число событий, при которых случайная величина $\alpha_t(\nu)$ равна 1. Учитывая наличие монотонной цепочки длины s с началом в t, монотонная цепочка длины $s+\nu-1$ имеет ν способов расположения на отрезке последовательности. Если зафиксировать положение монотонной цепочки длины $s+\nu-1$, то число таких цепочек равно $C_{s+\nu-1}$. Из определения $C_{s+\nu-1}$ следует, что остаются $2s-\nu-3$ элемента, которые могут принимать произвольные значения из множества $\{0, \ldots, N-1\}$. Из (14) получаем равенство (2) при $\nu \in \{1, \ldots, s-1\}$:

$$\lambda_\nu = \nu^{-1} \sum_{t=1}^{n} \frac{\nu C_{s+\nu-1} N^{2s-\nu-3}}{N^{3s-2}} = \frac{n C_{s+\nu-1}}{N^{s+\nu+1}}.$$

Для вычисления выражения (14) при $\nu \in \{s, \ldots, 2s-2\}$ подсчитаем число событий, при которых случайная величина $\alpha_t(\nu)$ равна 1. Всего $2s-2-\nu$ вариантов расположения монотонной цепочки длины $s+\nu-1$ на отрезке $(X_{t-s+1}, \ldots, X_{t+2s-2})$, когда она не начинается и не кончается на концах отрезка. Если зафиксировать положение монотонной цепочки, то число таких цепочек равно $C_{s+\nu-1}$. Если же монотонная цепочка длины $s+\nu-1$ начинается либо заканчивается на концах отрезка $(X_{t-s+1}, \ldots, X_{t+2s-2})$,

то в каждом из этих двух случаев число монотонных цепочек длины $s+\nu-1$ равно $B_{s+\nu-1}$. Из определения $B_{s+\nu-1}$ следует, что остаются $2s-\nu-2$ элемента, которые могут принимать произвольные значения из множества $\{0,\ldots,N-1\}$. Из (14) получаем равенство (2) при $\nu\in\{s,\ldots,2s-2\}$:

$$\lambda_\nu=\nu^{-1}\sum_{t=1}^{n}\frac{2B_{s+\nu-1}N^{2s-\nu-2}+(2s-2-\nu)\,C_{s+\nu-1}N^{2s-\nu-3}}{N^{3s-2}}=$$
$$=\frac{n\,(2B_{s+\nu-1}N+(2s-2-\nu)\,C_{s+\nu-1})}{\nu N^{s+\nu+1}}.$$

Наконец, вычислим выражение (14) при $\nu=2s-1$. В этом случае отрезок $(X_{t-s+1},\ldots,X_{t+2s-2})$ содержит монотонную цепочку длины $3s-2$. Число таких монотонных цепочек равно A_{2s-1}. Из (14) получаем равенство (2) при $\nu=2s-1$:

$$\lambda_{2s-1}=\frac{1}{2s-1}\sum_{t=1}^{n}\frac{A_{2s-1}}{N^{3s-2}}=\frac{nA_{2s-1}}{(2s-1)\,N^{3s-2}}.$$

На основе результатов теоремы 3 и с помощью выражений (2), (9), (11)–(13) имеем

$$d\left(L\left(\xi_n\left(s\right)\right),CP\left(\lambda_1,\lambda_2,\ldots\,\lambda_{2s-1},0,0,\ldots\right)\right)\leqslant$$
$$\leqslant c_1\left(\Lambda\right)\phi+c_2\left(\Lambda\right)\sum_{t=1}^{n}\left(\left(\mathbf{P}\left\{E_t\right\}\right)^2+\left(\mathbf{P}\left\{E_t\right\}\right)\left(\mathbf{E}\left(U_t+V_t\right)\right)+\mathbf{E}\left(I\left\{E_t\right\}V_t\right)\right)\leqslant$$
$$\leqslant\exp\left\{\sum_{k=1}^{\infty}\lambda_k\right\}\sum_{t=1}^{n}\left(1+4s-4+2s-2\right)\left(\left(\begin{array}{c}s+N-1\\s\end{array}\right)N^{-s}\right)^2=$$
$$=\exp\left\{\sum_{k=1}^{2s-1}\lambda_k\right\}\frac{n\,(6s-5)\,N^2}{(s+N)^2\,N^{2s}}\left(\left(\begin{array}{c}s+N\\s\end{array}\right)\right)^2.$$

Теорема 1 доказана.

Перейдём к доказательству теоремы 2. Рассмотрим в оценке (3) множитель

$$\exp\left\{\sum_{k=1}^{\infty}\lambda_k\right\}=$$
$$=\exp\left\{n\sum_{k=1}^{s-1}\frac{C_{s+k-1}}{N^{s+k+1}}+n\sum_{k=s}^{2s-2}\frac{(2B_{s+k-1}N+(2s-2-k)\,C_{s+k-1})}{kN^{s+k+1}}+\frac{nA_{2s-1}}{(2s-1)\,N^{3s-2}}\right\}\tag{15}$$

и покажем, что при переходе к пределу выражение (15) ограничено. Значит, требуется доказать, что существуют такие числа $M,n_0,s_0<\infty$, что для всех $n\geqslant n_0$, $s\geqslant s_0$ выполнено неравенство $\exp\left\{\sum_{k=1}^{2s-1}\lambda_k\right\}<M$.

Проверим это утверждение. Заметим, что найдётся такое число $M'<\infty$, при котором $\max\{A_s,B_s,C_s\}\leqslant M'\,(s+N)^{N-1}$. Заметим также, что при $k\in\{s,\ldots,2s-2\}$ найдётся такое число $M''<\infty$, что

$$\frac{2B_{s+k-1}N+(2s-2-k)\,C_{s+k-1}}{k}\leqslant M''\,(s+k-1+N)^{N-1}.$$

Кроме того, из условия (4) теоремы 2 следует, что существуют такие числа $n_0,s'<\infty$, что для всех $n\geqslant n_0$, $s\geqslant s'$ выполнено равенство $n\,(s+N)^{N-1}\,N^{-s-1}<\lambda$.

Тогда для каждого $k \in \{1, \ldots, 2s-1\}$ найдётся такое $M_1 < \infty$, что

$$\begin{aligned}\lambda_k \leqslant M_1 n \frac{(s+N+k)^{N-1}}{N^{s+k+1}} = M_1 n \frac{(s+N)^{N-1}}{N^{s+1}} \cdot \frac{(1+k/(s+N))^{N-1}}{N^k} < \\ < M_1 n \frac{(s+N)^{N-1}}{N^{s+1}} \left(\frac{e^{(N-1)/(s+N)}}{N}\right)^k < \lambda M_1 \left(\frac{e^{(N-1)/(s+N)}}{N}\right)^k.\end{aligned} \tag{16}$$

Выберем $s_0 = \max\{s', 2N\}$. Тогда

$$\frac{e^{(N-1)/(s+N)}}{N} \leqslant \frac{e^{(N-1)/(3N)}}{N} < \frac{e^{1/3}}{3}. \tag{17}$$

Подставив оценки (16) и (17) в (15), для любых $n \geqslant n_0$ и $s \geqslant s_0$ получим

$$\begin{aligned}\exp\left\{\sum_{k=1}^{2s-1} \lambda_k\right\} < \exp\left\{\lambda M_1 \sum_{k=1}^{2s-1} \left(\frac{e^{1/3}}{3}\right)^k\right\} < \exp\left\{\lambda M_1 \sum_{k=1}^{\infty} \left(\frac{e^{1/3}}{3}\right)^k\right\} = \\ = \exp\left\{\lambda M_1 \frac{e^{1/3}}{3 - e^{1/3}}\right\} < \infty.\end{aligned} \tag{18}$$

Применяя (3) и (18) при условиях теоремы 2, получаем

$$\begin{aligned}d\left(L\left(\xi_n\left(s\right)\right), CP\left(\lambda_1, \lambda_2, \ldots, \lambda_{2s-1}, 0, 0, \ldots\right)\right) \leqslant \\ \leqslant \exp\left\{\sum_{k=1}^{2s-1} \lambda_k\right\} \frac{n\left(6s-5\right)}{\left(s+N\right)^2 N^{2s-2}} \left(\left(\begin{array}{c} s+N \\ s \end{array}\right)\right)^2 = \mathrm{O}\left(\frac{ns\left(s+N\right)^{2N-2}}{N^{2s-2}}\right) = \mathrm{o}\left(1\right).\end{aligned} \tag{19}$$

Сформулируем лемму из [8] для оценки расстояния по вариации между двумя сложными распределениями Пуассона.

Лемма 1. Пусть $\Lambda^{(1)} = \left(\lambda_1^{(1)}, \lambda_2^{(1)}, \ldots\right)$ и $\Lambda^{(2)} = \left(\lambda_1^{(2)}, \lambda_2^{(2)}, \ldots\right)$, причём сходятся ряды $\sum\limits_{k=1}^{\infty} \lambda_k^{(1)} < \infty$ и $\sum\limits_{k=1}^{\infty} \lambda_k^{(2)} < \infty$. Тогда $d\left(CP\left(\Lambda^{(1)}\right), CP\left(\Lambda^{(2)}\right)\right) \leqslant \sum\limits_{k=1}^{\infty} \left|\lambda_k^{(1)} - \lambda_k^{(2)}\right|$.

Воспользуемся леммой 1 и оценим расстояние по вариации между распределениями $CP\left(\lambda_1, \lambda_2, \ldots, \lambda_{2s-1}, 0, 0, \ldots\right)$ и $CP\left(\lambda\left(1-N^{-1}\right), \lambda N^{-1}\left(1-N^{-1}\right), \lambda N^{-2}\left(1-N^{-1}\right), \ldots\right)$:

$$\begin{aligned}d\left(CP\left(\lambda_1, \lambda_2, \ldots, \lambda_{2s-1}, \right.\right. & \left.\left. 0, 0, \ldots\right), CP\left(\lambda\left(1-N^{-1}\right), \lambda N^{-1}\left(1-N^{-1}\right), \ldots\right)\right) \leqslant \\ & \leqslant \sum_{k=1}^{\infty} \left|\lambda_k - \lambda N^{-k+1}\left(1-N^{-1}\right)\right|.\end{aligned} \tag{20}$$

Теперь докажем, что сумма в правой части (20) стремится к 0 при условиях теоремы 2. Зададим произвольное малое положительное число $\varepsilon > 0$ и выберем некоторое натуральное число $k'\left(\varepsilon\right)$, удовлетворяющее условиям

$$\sum_{k=k'(\varepsilon)+1}^{\infty} \lambda N^{-k+1}\left(1-N^{-1}\right) < \frac{\varepsilon}{3}; \tag{21}$$

$$\sum_{k=k'(\varepsilon)+1}^{\infty} \lambda_k < \frac{\varepsilon}{3}. \tag{22}$$

Условие (21) выполнить легко: надо взять достаточно большое $k'(\varepsilon)$. Докажем условие (22). Из (16) и (17) следует, что существуют такие числа n_0, $s_0 < \infty$, что для всех $n \geqslant n_0$, $s \geqslant s_0$ выполнено неравенство

$$\lambda_k < \lambda M_1 \left(\frac{e^{1/3}}{3}\right)^k.$$

Значит,

$$\sum_{k=1}^{2s-1} k\lambda_k < \lambda M_1 \sum_{k=1}^{2s-1} \left(\frac{e^{1/3}}{3}\right)^k < \lambda M_1 \sum_{k=1}^{\infty} \left(\frac{e^{1/3}}{3}\right)^k = \frac{27\lambda M_1}{e^{1/3}(1-e^{1/3})^2} \equiv M_2. \tag{23}$$

Для выполнения условия (22) воспользуемся соотношением (23):

$$\sum_{k=k'(\varepsilon)+1}^{\infty} \lambda_k < \frac{1}{k'(\varepsilon)} \sum_{k=1}^{2s-1} k\lambda_k < \frac{M_2}{k'(\varepsilon)}.$$

Следовательно, взяв достаточно большое число $k'(\varepsilon)$, получим выполнение (22). Осталось заметить, что при условиях теоремы 2

$$\sum_{k=1}^{k'(\varepsilon)} \left|\lambda_k - \lambda N^{-k+1}\left(1-N^{-1}\right)\right| \leqslant \sum_{k=1}^{k'(\varepsilon)} \lambda_k + \sum_{k=1}^{k'(\varepsilon)} \lambda N^{-k+1}\left(1-N^{-1}\right) \to 0$$

как сумма фиксированного числа величин, стремящихся к нулю. Значит, начиная с некоторого момента,

$$\sum_{k=1}^{k'(\varepsilon)} \left|\lambda_k - \lambda N^{-k+1}\left(1-N^{-1}\right)\right| < \frac{\varepsilon}{3}. \tag{24}$$

Из (21), (22) и (24) следует, что, начиная с некоторого момента,

$$\sum_{k=1}^{\infty} \left|\lambda_k - \lambda N^{-k+1}\left(1-N^{-1}\right)\right| \leqslant$$

$$\leqslant \sum_{k=1}^{k'(\varepsilon)} \left|\lambda_k - \lambda N^{-k+1}\left(1-N^{-1}\right)\right| + \sum_{k=k'(\varepsilon)+1}^{\infty} \lambda_k + \sum_{k=k'(\varepsilon)+1}^{\infty} \lambda N^{-k+1}\left(1-N^{-1}\right) < \varepsilon.$$

В силу произвольности выбора $\varepsilon > 0$ это означает, что

$$\sum_{k=1}^{\infty} \left|\lambda_k - \lambda N^{-k+1}\left(1-N^{-1}\right)\right| \to 0. \tag{25}$$

Используя (20) и (25), получаем, что при условиях теоремы 2

$$d\left(CP\left(\lambda_1, \lambda_2, \ldots, \lambda_{2s-1}, 0, 0, \ldots\right), CP\left(\lambda\left(1-N^{-1}\right), \lambda N^{-1}\left(1-N^{-1}\right), \ldots\right)\right) \to 0. \tag{26}$$

Наконец, из (19) и (26) в силу неравенства треугольника для расстояния по вариации при условиях теоремы 2 следует, что

$$d\left(L\left(\xi_n(s)\right), CP\left(\lambda\left(1-N^{-1}\right), \lambda N^{-1}\left(1-N^{-1}\right), \lambda N^{-2}\left(1-N^{-1}\right), \ldots\right)\right) \to 0,$$

а значит, следует и сходимость $L\left(\xi_n(s)\right)$ к сложному пуассоновскому распределению $CP\left(\lambda\left(1-N^{-1}\right), \lambda N^{-1}\left(1-N^{-1}\right), \lambda N^{-2}\left(1-N^{-1}\right), \ldots\right)$.

Теорема 2 доказана.

ЛИТЕРАТУРА

1. *Гончаров В. Л.* Из области комбинаторики // Изв. АН СССР. Сер. матем. 1944. Т. 8. Вып. 1. С. 3–48.
2. *Wolfowitz J.* Asymptotics distribution of runs up and down // Ann. Math. Statist. 1944. V. 15. P. 163–172.
3. *David F. N. and Barton D. E.* Combinatorial Chance. N.Y.: Hafner Publishing Co., 1962.
4. *Pittel B. G.* Limiting behavior of a process of runs // Ann. Probab. 1981. V. 9. No. 1. P. 119–129.
5. *Chryssaphinou O., Papastavridis S., and Vaggelatou E.* Poisson approximation for the non-overlapping appearances of several words in Markov chains // Combinatorics, Probability and Computing. 2001. V. 10. No. 4. P. 293–308.
6. *Меженная Н. М.* Многомерная нормальная теорема для числа монотонных серий заданной длины в равновероятной случайной последовательности // Обозр. прикл. промышл. матем. 2007. Т. 14. Вып. 3. С. 503–505.
7. *Roos V.* Stein's method for compound Poisson approximation: the local approach // Ann. Appl. Probab. 1994. V. 4. No. 4. P. 1177–1187.
8. *Bollobas B, Janson S, and Riordan O.* Sparse random graphs with clustering // Random Structures and Algorithms. 2011. V. 38. P. 269–323.

REFERENCES

1. *Goncharov V. L.* Iz oblasti kombinatoriki [From the combinatorics]. Proc. of the Academy of Sciences USSR, Ser. Math., 1944, vol. 8, iss. 1, pp. 3–48. (in Russian)
2. *Wolfowitz J.* Asymptotics distribution of runs up and down. Ann. Math. Statist., 1944, vol. 15, pp. 163–172.
3. *David F. N. and Barton D. E.* Combinatorial Chance. N.Y., Hafner Publishing Co., 1962.
4. *Pittel B. G.* Limiting behavior of a process of runs. Ann. Probab., 1981, vol. 9, no. 1, pp. 119–129.
5. *Chryssaphinou O., Papastavridis S., and Vaggelatou E.* Poisson approximation for the non-overlapping appearances of several words in Markov chains. Combinatorics, Probability and Computing, 2001, vol. 10, no. 4, pp. 293–308.
6. *Mezhennaya N. M.* Mnogomernaya normal'naya teorema dlya chisla monotonnykh seriy zadannoy dliny v ravnoveroyatnoy sluchaynoy posledovatel'nosti [Multivariate normal theorem for the number of monotonous series of predetermined length in an equiprobable random sequence]. Obozr. Prikl. Promyshl. Matem., 2007, vol. 14, iss. 3, pp. 503–505. (in Russian)
7. *Roos V.* Stein's method for compound Poisson approximation: the local approach. Ann. Appl. Probab., 1994, vol. 4, no. 4, pp. 1177–1187.
8. *Bollobas B, Janson S, and Riordan O.* Sparse random graphs with clustering. Random Structures and Algorithms, 2011, vol. 38, pp. 269–323.

УДК 512.55

ПОЧТИ НИЛЬПОТЕНТНЫЕ МНОГООБРАЗИЯ АЛГЕБР ЛЕЙБНИЦА

Ю. Ю. Фролова, О. В. Шулежко

Ульяновский государственный университет, г. Ульяновск, Россия

Представлен новый результат, касающийся многообразий алгебр Лейбница. Доказано, что в случае нулевой характеристики основного поля существует ровно два почти нильпотентных многообразия алгебр Лейбница. Доказательство носит комбинаторный характер.

Ключевые слова: *алгебра Ли, алгебра Лейбница, почти нильпотентное многообразие, диаграмма Юнга.*

DOI 10.17223/20710410/28/3

ALMOST NILPOTENT VARIETIES OF LEIBNIZ ALGEBRAS

Yu. Yu. Frolova, O. V. Shulezhko

Ulyanovsk State University, Ulyanovsk, Russia

E-mail: yuyufrolova@mail.ru, ol.shulezhko@gmail.com

It is proved that there exist exactly two almost nilpotent varieties of Leibniz algebras over a field of zero characteristic.

Keywords: *Leibniz algebra, Lie algebra, almost nilpotent variety, Young diagram.*

На протяжении всей работы характеристика основного поля Φ равна нулю. Напомним, что векторное пространство над полем Φ называется линейной алгеброй, если на нём задана бинарная билинейная операция. Класс линейных алгебр, в которых выполняется некоторый фиксированный набор тождественных соотношений, называется многообразием. В работе не предполагается ассоциативность, поэтому при более чем двух сомножителях имеет значение расположение скобок. Более того, в случае алгебр Лейбница любое произведение представимо в виде линейной комбинации левонормированных элементов. Договоримся опускать скобки в случае их левонормированной расстановки, то есть $abc = (ab)c$. Все неопределённые в данной работе понятия можно найти в [1, 2].

Хорошо известно, что в случае нулевой характеристики основного поля любое тождество эквивалентно некоторой системе полилинейных тождеств, и исследование строения полилинейных частей относительно свободной алгебры многообразия дает полную информацию об этом многообразии.

В свободной алгебре многообразия V со счётным множеством свободных образующих $X = \{x_1, x_2, \ldots\}$ рассмотрим множество полилинейных элементов степени n от $x_1, x_2, \ldots x_n$. Они образуют векторное пространство $P_n(V)$, называемое полилинейной компонентой относительно свободной алгебры, размерность которого обозначим $c_n(V)$, $n = 1, 2, \ldots$ Известно, что полилинейную компоненту степени n можно рассматривать как модуль над групповым кольцом ΦS_n симметрической группы S_n.

Для этого определим действие перестановки σ следующим образом: $\sigma(x_{i_1}x_{i_2}\dots x_{i_n}) = x_{\sigma(i_1)}x_{\sigma(i_2)}\dots x_{\sigma(i_n)} \in S_n$. Это действие симметрической группы S_n превращает полилинейную компоненту P_n в левый ΦS_n-модуль.

Известно, что с точностью до изоморфизма неприводимые ΦS_n-модули можно описывать на языке разбиений и диаграмм Юнга. Разбиением числа n называют набор целых положительных чисел $\lambda = (\lambda_1, \dots, \lambda_k)$, при этом $\lambda_1 \geqslant \dots \geqslant \lambda_k > 0$ и $n = \sum\limits_{i=1}^{k} \lambda_i$. Разбиение λ числа n обозначают $\lambda \vdash n$. Для каждого такого разбиения λ строится диаграмма Юнга, состоящая из k строк, причём строка с номером i содержит λ_i клеток. Если диаграмму Юнга, соответствующую разбиению λ числа n, заполнить числами от 1 до n, то получим таблицу Юнга, соответствующую данной диаграмме.

Так как характеристика основного поля равна нулю, по теореме Машке полилинейную часть степени n можно разложить в прямую сумму неприводимых подмодулей. Строение модуля $P_n(V)$ можно представить на «языке характеров». Рассмотрим разложение характера модуля $P_n(V)$ в целочисленную комбинацию неприводимых характеров:

$$\chi_n(V) = \chi(P_n(V)) = \sum_{\lambda \vdash n} m_\lambda \chi_\lambda. \tag{1}$$

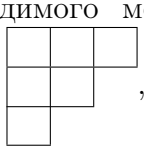
Здесь m_λ — кратность неприводимого характера χ_λ, отвечающего разбиению λ. Асимптотическое поведение размерности $c_n(V)$ пространства $P_n(V)$ определяет рост многообразия. Число слагаемых $l_n(V) = \sum\limits_{\lambda \vdash n} m_\lambda$ в сумме (1) называют кодлиной многообразия.

Размерность неприводимого ΦS_n-модуля, построенного по некоторому разбиению λ, которую будем обозначать d_λ, определяется формулой «крюков» (см., например, [1, с. 113])

$$d_\lambda = \frac{n!}{\prod\limits_{i,j \in d} |h_{ij}|},$$

где длина крюка $|h_{ij}| = (\lambda_i - i) + (\lambda'_j - j) + 1$; λ_i — длина i-й строки; λ'_j — длина j-го столбца диаграммы d.

Например, размерность неприводимого модуля, соответствующего диаграмме



равна $d_{(3,2,1)} = \dfrac{6!}{1 \cdot 3 \cdot 5 \cdot 1 \cdot 3 \cdot 1} = 16$.

Будем говорить, что многообразие V является почти нильпотентным, если V не нильпотентно, но каждое собственное подмногообразие W многообразия V нильпотентно.

Рассмотрим состояние вопроса описания почти нильпотентных многообразий в классах ассоциативных алгебр и алгебр Ли.

Известно, что в классе ассоциативных алгебр единственным почти нильпотентным многообразием является многообразие всех ассоциативно-коммутативных алгебр AC. По тождеству ассоциативности скобки в полилинейных элементах можно опускать, в силу тождества коммутативности образующие в произведениях можно менять местами. Поэтому полилинейная часть данного многообразия является одномерным пространством ($c_n(P_n(AC)) = 1$), базис которого — любой моном, например $x_1x_2\dots x_n$. Разложение характера имеет вид $\chi_n(AC) = \chi_{(n)}$, кодлина равна 1, то есть $l_n(AC) = 1$ для любого n. Понятно, что тождество $x_1^n \equiv 0$ соответствует разбиению (n). Так как

поле Ф лежит в многообразии AC, само многообразие AC не является нильпотентным. Однако если рассмотреть любое собственное подмногообразие многообразия всех ассоциативно-коммутативных алгебр AC, то в нём должно выполняться тождество, невыполнимое в AC. Значит, в этом подмногообразии выполняется тождество $x^n \equiv 0$. Тогда по теореме Нагаты — Хигмана [3, с. 152–153] получаем, что оно является нильпотентным. Не претендуя на авторство, сформулируем следующее утверждение.

Теорема 1. Пусть W — почти нильпотентное ассоциативное многообразие. Тогда $W = AC$.

Доказательство. Пусть многообразие W является почти нильпотентным, но не совпадает с многообразием AC. Понятно, что ненильпотентное многообразие AC не является собственным подмногообразием многообразия W. Отсюда следует, что в многообразии W выполняется тождество $x^n \equiv 0$. По теореме Нагаты — Хигмана получаем, что W является нильпотентным. ■

Напомним, что алгеброй Ли называется алгебра, в которой выполнено тождество антикоммутативности $x^2 \equiv 0$ и тождество Якоби $(xy)z + (yz)x + (zx)y \equiv 0$. Введём обозначение Y для оператора умножения справа на элемент y; запишем $aY = ay$. Тогда $x\underbrace{y \dots y}_{m} = xY^m$.

Рассмотрим многообразие метабелевых алгебр Ли, которое обозначим A^2. Это многообразие определяется тождеством

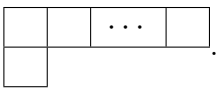
$$(x_1x_2)(x_3x_4) \equiv 0. \tag{2}$$

Из тождеств антикоммутативности и Якоби следует, что в многообразии A^2 выполняется следующее тождество:

$$x_1(x_2x_3) = x_1x_2x_3 - x_1x_3x_2. \tag{3}$$

Из (2) и (3) получаем, что в элементах пространства $P_n(A^2)$ образующие, начиная с третьего, можно менять местами: $x_1x_2x_4x_3 \equiv x_1x_2x_3x_4$. Более подробно данное многообразие описано в [1]. Кроме того, в алгебре Ли любой элемент в пространстве P_n можно представить в виде линейной комбинации элементов вида $x_nx_{j_1}x_{j_2} \dots x_{j_{(n-1)}}$. Значит, полилинейная часть $P_n(A^2)$ является линейной оболочкой элементов $x_nx_ix_{i_1} \dots x_{i_{(n-2)}}$, где $i_1 < i_2 < \dots < i_{n-2}$.

Размерность полилинейной части многообразия A^2 равна $n-1$, то есть $c_n(A^2) = n-1$. Разложение характера неприводимого модуля этого многообразия имеет вид $\chi_n(A^2) = \chi_{(n-1,1)}$, а кодлина многообразия равна единице: $l_n(A^2) = 1$.

Соответствующая разбиению $(n-1, 1)$ диаграмма Юнга имеет вид  .

Такой диаграмме соответствует тождество $x_1x_2{X_1}^{n-2} - x_2x_1{X_1}^{n-2} \equiv 0$.

Воспользовавшись свойством антикоммутативности, получим, что это тождество эквивалентно так называемому тождеству энгелевости

$${x_2X_1}^{n-1} \equiv 0. \tag{4}$$

Заметим, что в случае многообразия алгебр Ли тождество энгелевости имеет вид $xY^m \equiv 0$, а не $xX^m \equiv 0$ ввиду антикоммутативности $x^2 \equiv 0$. Тождество (4) не выполняется в многообразии A^2, так как в нём лежит метабелева ненильпотентная алгебра с базисом e, h и таблицей умножения $he = -eh = h$, $ee = hh = 0$ [1, с. 173]. Изложим в виде теоремы хорошо известный факт.

Теорема 2. Пусть W — почти нильпотентное многообразие алгебр Ли. Тогда $W = A^2$.

Доказательство. Пусть многообразие W является почти нильпотентным, но не совпадает с многообразием A^2. Понятно, что ненильпотентное многообразие A^2 не является собственным подмногообразием многообразия W. Так как кратность $m_{(n-1,1)} = 1$ как в самом многообразии A^2, так и в многообразии всех алгебр Ли, то в W выполняется тождество (4). Тогда из результата Е. И. Зельманова [4] получаем, что в W выполняется тождество нильпотентности $x_1x_2\dots x_m \equiv 0$ для некоторого натурального числа m. Значит, W — нильпотентное многообразие. Таким образом, получили, что A^2 — единственное почти нильпотентное многообразие алгебр Ли. ■

Перейдём к изложению основной части работы: к описанию почти нильпотентных многообразий в классе алгебр Лейбница. Напомним, что алгеброй Лейбница над полем Ф называется линейная алгебра, удовлетворяющая тождеству Лейбница

$$(xy)z \equiv (xz)y + x(yz).$$

В другом виде получаем

$$x(yz) \equiv xyz - xzy. \tag{5}$$

Отметим, что если в алгебре Лейбница выполняется тождество антикоммутативности $x^2 \equiv 0$, то (5) эквивалентно тождеству Якоби. Значит, любая алгебра Ли является алгеброй Лейбница, и многообразие алгебр Ли A^2 также является почти нильпотентным многообразием алгебр Лейбница.

Из тождества (5) следует, что любой элемент может быть представлен в виде линейной комбинации левонормированных элементов. Поэтому определение нильпотентной алгебры следующее: алгебра называется нильпотентной алгеброй ступени c, если в ней выполняется тождество $x_1x_2\dots x_{c+1} \equiv 0$ с левонормированной расстановкой скобок, но не выполняется тождество $x_1x_2\dots x_c \equiv 0$. В многообразии N_c нильпотентных алгебр Лейбница ступени c коразмерность $c_n(N_c)$ равна нулю при $n > c$. Рассмотрим многообразие алгебр Лейбница ${}_2N$, в котором выполнено тождество $x(yz) \equiv 0$. Полное описание этого многообразия приведено в [5]. Для многообразия ${}_2N$ справедливо следующее разложение для характера: $\chi_n({}_2N) = \chi_{(n)} + \chi_{(n-1,1)}$.

Кроме того, в работе [5] получено, что соответствующие неприводимые модули порождаются полной линеаризацией следующих элементов:

$$g_{(n)} = x_1X_1^{n-1}, \quad g_{(n-1,1)} = x_2x_1X_1^{n-2} - x_1x_2X_1^{n-2}.$$

Пусть A — некоторая алгебра Лейбница. В работе [5] показано, что линейная оболочка квадратов элементов алгебры является идеалом, который обозначим I.

Утверждение 1. Для любого элемента b из алгебры A и любого элемента c из идеала I выполняется равенство $bc = 0$, то есть идеал является правым аннулятором алгебры. В частности, идеал I является алгеброй с нулевым умножением.

Доказательство. Элемент вида $ab + ba$ принадлежит множеству I, так как $ab + ba = (a + b)^2 - a^2 - b^2$. Пусть b — элемент алгебры A, тогда произведение $a^2b = (ab)a + a(ab)$ принадлежит множеству I. Поэтому I является правым идеалом. Из тождества $b(aa) \equiv 0$ следует, что произведение ba^2 является нулевым и поэтому также принадлежит идеалу.

Рассмотрим произведение $(a+I)(a+I) = a^2 + I = I$. Значит, в фактор-алгебре A/I выполняется тождество антикоммутативности, по модулю которого тождество (3) эквивалентно тождеству Якоби, то есть фактор-алгебра A/I является алгеброй Ли. ■

В работе [6, с. 39 (следствие 2.3)] доказано, что в классе разрешимых алгебр Лейбница существует ровно два описанных выше почти нильпотентных многообразия. Основной целью данной работы является обобщение этого результата на случай всех алгебр Лейбница.

Теорема 3. В случае нулевой характеристики основного поля существует ровно два почти нильпотентных многообразия алгебр Лейбница. Это многообразие метабелевых алгебр Ли A^2 и многообразие левонильпотентных ступени не выше двух алгебр Лейбница ${}_2N$.

Доказательство. Пусть U — новое почти нильпотентное многообразие. Предположим, что все тождества, которые выполнены в U, выполняются и в ${}_2N$. Тогда ${}_2N \subseteq U$, но многообразие U почти нильпотентно, а значит, ${}_2N = U$. Таким образом, в U выполнены какие-то тождества, которые не выполняются в ${}_2N$. Возможны два случая, отвечающие разбиениям (n) и $(n-1,1)$.

Тождество, соответствующее разбиению $(n) \vdash n$, имеет вид $x_1 X_1^{n-1} \equiv 0$.

Разбиению $(n-1,1) \vdash n$ соответствуют тождества вида

$$\sum_{s=0}^{n-2} \alpha_s \overline{x_1} X_1^s \overline{x_2} X_1^{n-s-2} \equiv 0. \tag{6}$$

Левая часть тождества является линейной комбинацией элементов, соответствующих стандартным таблицам Юнга. Стандартная таблица для (6) выглядит следующим образом:

1	2	$\cdots$	$\widehat{s+2}$	$\cdots$	n
$s+2$					

Соответствующий этой таблице элемент имеет вид $\overline{x_1} X_1^s \overline{x_2} X_1^{n-s-2}$. Более подробно о теории представлений симметрической группы можно найти в [2].

Пусть в (6) $\alpha = \sum \alpha_s = 0$. В многообразии ${}_2N$, согласно тождествам $x(yz) \equiv \equiv xyz - xzy$ и $x(yz) \equiv 0$, сомножители, начиная со второго, можно менять местами. Поэтому (6) по модулю тождеств многообразия ${}_2N$ имеет вид $\alpha \overline{x_1}\,\overline{x_2} X_1^{n-2} \equiv 0$. Таким образом, оно выполняется в этом многообразии вопреки предположению. Получили противоречие, следовательно, $\alpha = \sum \alpha_s \neq 0$. Подставив в это тождество вместо x_2 элемент x_1^2, получим $-\alpha x_1 X_1{}^n \equiv 0$.

Итак, в обоих рассматриваемых случаях получили, что в многообразии U выполнено тождество $x_1 X_1^m \equiv 0$ для некоторого натурального m ($m = n-1$ или $m = n$).

Подставим в последнее тождество вместо x_1 сумму $y + x^2$. Так как все слагаемые вида $y \ldots (x^2) \ldots$ равны нулю, в качестве следствия получаем тождество вида $x^2 Y^m \equiv 0$ или эквивалентные ему частичные линеаризации:

$$xyY^m \equiv -yxY^m; \tag{7}$$

$$x^2 f(Y_1, \ldots, Y_m) \equiv 0, f(Y_1, \ldots, Y_m) = \sum_{p \in S_m} Y_{p_{(1)}} \ldots Y_{p_{(m)}}. \tag{8}$$

Рассмотрим алгебру Лейбница L из многообразия U и множество элементов

$$I_f = \{a \in L : \ af(Y_1, \ldots, Y_m) = 0 \ \text{ для любых } \ y_1, \ldots, y_m \in L\}.$$

Докажем, что I_f является правым идеалом.

Пусть $a \in I_f$, значит, $af(Y_1, \ldots, Y_m) = 0$ для любых $y_1, \ldots, y_m \in L$. Умножим на $b \in L$ справа: $af(Y_1, \ldots, Y_m)b = 0$, но, с другой стороны, дифференцируя при помощи элемента b по тождеству Лейбница, получим

$$abf(Y_1, \ldots, Y_m) + \sum_{s=1}^{m} af(Y_1, \ldots, [Y_s, B], \ldots, Y_m) = 0,$$

где $[Y_s, B] = Y_sB - BY_s$ — коммутатор линейных операторов.

По определению множества I_f каждое слагаемое суммы отдельно равно нулю. В качестве примера рассмотрим случай $m = 2$. Тогда многочлен имеет вид $f(Y_1, Y_2) = Y_1Y_2 + Y_2Y_1$, получаем, что $ay_1y_2 + ay_2y_1 = 0$. Умножим последнее равенство на b справа и, используя правило дифференцирования, получим

$$aby_1y_2 + a(y_1b)y_2 + ay_1(y_2b) + aby_2y_1 + a(y_2b)y_1 + ay_2(y_1b) = \\ = abf(Y_1, Y_2) + af([Y_1, B], Y_2) + af(Y_1, [Y_2, B]) = 0.$$

Так как два последних слагаемых по определению множества I_f равны нулю, получаем, что $abf(Y_1, Y_2) = 0$ для любых элементов y_1, y_2 из алгебры L.

В предыдущих выкладках мы использовали тот факт, что в алгебрах Лейбница выполняется равенство $x(yz) = ayz - azy = a[Y, Z]$. Таким образом, $abf(Y_1, \ldots, Y_m) = 0$ для любых $y_s \in L$. Следовательно, $ab \in I_f$.

Из тождества (7) следует, что $abX^m = -baX^m \equiv 0$, а значит, произведение ba принадлежит идеалу I_f и, следовательно, идеал является двусторонним. Заметим, что в идеале I_f справедливо тождество $zf(Y_1, \ldots, Y_m) \equiv 0$, которое эквивалентно тождеству $zX^m \equiv 0$, то есть выполняется энгелевость, и по теореме из работы [7] идеал I_f является нильпотентным. Из тождества (8) следует, что для любого элемента a из алгебры L квадрат этого элемента a^2 принадлежит идеалу I_f, следовательно, в факторалгебре L/I_f выполняется тождество $x^2 \equiv 0$, а значит, L/I_f является алгеброй Ли.

Рассмотрим два случая:

1) Фактор-алгебра L/I_f не является нильпотентной алгеброй Ли. В этом случае, так как L/I_f принадлежит многообразию U, A^2 является ненильпотентным подмногообразием U. По теореме 2 в этом случае $U = A^2$.
2) Фактор-алгебра L/I_f является нильпотентной некоторой ступени c. По определению идеала I_f в алгебре L выполняется тождество $y_1y_2 \ldots y_{c+1}X^m \equiv 0$. Следовательно, выполняется также тождество энгелевости, и по теореме из работы [7] многообразие U является нильпотентным. Это противоречит условию теоремы.

Таким образом, ${}_2N$ и A^2 — единственные почти нильпотентные многообразия алгебр Лейбница. ■

Авторы выражают благодарность С. П. Мищенко за ценные советы и постоянное внимание к работе.

ЛИТЕРАТУРА

1. *Бахтурин Ю. А.* Тождества в алгебрах Ли. М.: Наука, 1985. 448 с.
2. *Giambruno A. and Zaicev M.* Polynomial identities and asymptotic methods. American Mathematical Society Providence, RI, 2005. 352 p.
3. *Ширшов А. И. и др.* Кольца, близкие к ассоциативным. М.: Наука, 1978. 432 с.
4. *Зельманов Е. И.* Об энгелевых алгебрах Ли // ДАН СССР. 1987. Т. 292. № 2. С. 265–268.

5. *Drensky V. and Cattaneo G. M. P.* Varieties of metabelian Leibniz algebras // J. Algebra Appl. 2002. V. 1. No. 1. P. 31–50.
6. *Черевatенко О. И.* Некоторые эффекты роста тождеств линейных алгебр: дис. ... канд. физ.-мат. наук. Ульяновск, 2008. 69 с.
7. *Фролова Ю. Ю.* О нильпотентности энгелевой алгебры Лейбница // Вестник Московского государственного университета. Сер. 1. Математика. Механика. 2011. № 3. С. 63–65.

REFERENCES

1. *Bakhturin Yu. A.* Tozhdestva v algebrakh Li [Identities of Lie Algebras]. Moscow, Nauka Publ., 1985. 448 p. (in Russian)
2. *Giambruno A. and Zaicev M.* Polynomial identities and asymptotic methods. American Mathematical Society Providence, RI, 2005. 352 p.
3. *Shirshov A. I. et al.* Kol'tsa, blizkie k assotsiativnym [Rings that are Nearly Associative]. Moscow, Nauka Publ., 1978. 432 p. (in Russian)
4. *Zel'manov E. I.* Ob engelevykh algebrakh Li [On Engel Lie algebras]. Reports USSR Academy of Sciences, 1987, vol. 292, no. 2, pp. 265–268. (in Russian)
5. *Drensky V. and Cattaneo G. M. P.* Varieties of metabelian Leibniz algebras. J. Algebra Appl., 2002, vol. 1, no. 1, pp. 31–50.
6. *Cherevatenko O. I.* Nekotorye effekty rosta tozhdestv lineynykh algebr [Some Effects of Linear Algebra Identities Increase.] PhD Thesis, Ul'yanovsk, 2008. 69 p. (in Russian)
7. *Frolova Yu. Yu.* O nil'potentnosti engelevoy algebry Leybnitsa [On the nilpotency of Engel Leibniz algebra.] MSU Bulletin. Ser. 1. Mathematics. Mechanics. 2011, no. 3, pp. 63–65. (in Russian)

УДК 512.55

КРИТЕРИИ ФУНКЦИОНАЛЬНОЙ РАЗДЕЛИМОСТИ КВАДРАТИЧНЫХ БУЛЕВЫХ ПОРОГОВЫХ ФУНКЦИЙ

А. Н. Шурупов

Московский государственный технический университет радиотехники, электроники и автоматики (МИРЭА), г. Москва, Россия

Работа продолжает исследование функциональной структуры булевых функций, задаваемых действительными линейными неравенствами. Рассматриваются булевы функции, определяемые одним нелинейным неравенством второй степени. Многочлены второй степени среди всех нелинейных многочленов обладают наименьшим размером задания, т. е. свойством, существенным в ряде прикладных задач. Доказаны три критерия функциональной разделимости для булевых квадратичных пороговых функций. Второй критерий не требует анализа табличного задания функции и формулируется в терминах пороговой структуры.

Ключевые слова: *функциональная разделимость, декомпозиция, булевы пороговые функции, квадратичные неравенства.*

DOI 10.17223/20710410/28/4

FUNCTIONAL DECOMPOSABILITY CRITERIA FOR QUADRATIC THRESHOLD BOOLEAN FUNCTIONS

A. N. Shurupov

Moscow State Institute of Radio Engineering, Electronics and Automation, Moscow, Russia

E-mail: ashurupov@mail.ru

Threshold functions provide a simple but fundamental model for many questions investigated in image recognition, artificial neural networks and many other areas. In this paper, the results in Boolean threshold function decomposition are advanced to Boolean functions represented by one quadratic inequality. Quadratic polynomials are the most compact non-linear polynomials and this property sometimes is quite important. We prove three criteria for non-trivial decomposition of quadratic Boolean threshold functions. One of them can be applied without analysis of truth table and only uses the threshold structure parameters.

Keywords: *Boolean functions, threshold functions, decomposition, quadratic inequalities.*

1. Основные понятия

Проблема описания функционально разделимых булевых функций, то есть функций, допускающих бесповторную декомпозицию, исследуется много лет. Прекрасный обзор и ряд глубоких результатов по декомпозиции k-значных функций содержится в [1]. В настоящей работе рассматривается функциональная разделимость булевых функций из одного специфического класса — булевых пороговых функций, задаваемых квадратичными неравенствами. Интерес к пороговым функциям в настоящее

время обуславливается их применениями для решения задач распознавания образов, в искусственных нейронных сетях и других областях [2]. Вопросы функциональной разделимости для булевых (линейных) пороговых функций рассмотрены в [3].

В дальнейшем потребуются следующие понятия (см. [1, 4–6]).

Пусть $F_2(n) = \{f : \{0,1\}^n \to \{0,1\}\}$ — множество всех булевых функций от n переменных.

Определение 1. Пусть K — некоторый непустой класс булевых функций. Если для функции $f \in F_2(n)$ существует представление

$$f(x_1, \ldots, x_n) = \varphi\big(\psi(x_{i_1}, \ldots, x_{i_m}), x_{i_{m+1}}, \ldots, x_{i_n}\big)$$

для некоторых m, перестановки индексов переменных $(i_1, \ldots, i_n)$ и функций $\varphi, \psi \in K$, то говорят, что f допускает *простую бесповторную декомпозицию в классе* K. При этом переменные $x_{i_1}, \ldots, x_{i_m}$ называются *связанными*, а $x_{i_{m+1}}, \ldots, x_{i_n}$ — *свободными*. Везде далее в силу рассмотрения только бесповторных декомпозиций будем опускать термин «бесповторная». Простая декомпозиция называется *нетривиальной*, если $1 < m < n$. Функции, допускающие нетривиальные простые декомпозиции, называются *функционально разделимыми* над соответствующим классом функций.

Заметим, что определение 1 является более общим по отношению к определению 2.15 работы [6], так как рассматривает декомпозицию в произвольном классе булевых функций. Кроме того, в смысле определения 2.15 указанной работы функционально неразделимой будет булева функция $f(x_1, \ldots, x_4) = (x_1x_2 \oplus x_1x_3 \oplus x_2x_3)x_4$, в то время как она является очевидно разложимой (и разделимой в смысле определения 1).

Далее без особых оговорок будем в случаях, ясных по контексту изложения, использовать для неотрицательных целых чисел и векторов — их двоичных представлений — одни и те же обозначения. *Таблица разбиения* $T^{(m)}$ *булевой функции* $f(x_1, \ldots, x_n)$ *с параметром* (*разбиения*) m определяется как матрица, состоящая из 2^m строк и 2^{n-m} столбцов с элементами $t_{u,v} = f(u, v) = f(u_1, \ldots, u_m, v_1, \ldots, v_{n-m})$. Известен [4] критерий функциональной разделимости, приводимый здесь для удобства в формулировке [3].

Теорема 1. Булева функция f допускает простую декомпозицию вида

$$f(x_1, \ldots, x_n) = \varphi\big(\psi(x_1, \ldots, x_m), x_{m+1}, \ldots, x_n\big) \tag{1}$$

в том и только в том случае, когда в $T^{(m)}$ существует не более двух типов различных строк.

Введём определение булевых функций, задаваемых квадратичными неравенствами, функциональные свойства которых изучаются в этой работе. Более общее определение полиномиально-задаваемой (или полиномиальной, как в [7]) пороговой функции можно найти вместе с богатой библиографией и обзором результатов по пороговым функциям в [8]. Эти понятия приводятся также в [7], как и следующее определение.

Определение 2. Булева функция $f \in F_2(n)$ называется *квадратичной пороговой* (*к.п.б.ф.*), если существует многочлен $q \in \mathbb{R}[x_1, \ldots, x_n]$, такой, что

$$\deg q(x_1, \ldots, x_n) = 2; \tag{2}$$

$$f(x_1, \ldots, x_n) = 0 \Leftrightarrow q(x_1, \ldots, x_n) \leqslant 0. \tag{3}$$

Традиционные пороговые функции в смысле определения 2 могут быть названы линейными пороговыми функциями. Отметим, что хотя в [7] булева функция в смысле определения 2 называется знаковой функцией многочлена q, а сам многочлен q — пороговым элементом, мы будем использовать также терминологию определения 2 для сохранения преемственности с теорией булевых линейных пороговых функций.

Многочлен q из определения 2 может быть представлен в виде $q(x_1, \ldots, x_n) = l(x_1, \ldots, x_n) + q_2(x_1, \ldots, x_n)$, где l — линейная часть многочлена, а q_2 — квадратичная часть. Так как для булевых переменных справедливо равенство $x = x^2$, многочлен $l = a_1x_1 + \ldots + a_nx_n + a_0$ совпадает с $q_1 = a_1x_1^2 + \ldots + a_nx_n^2 + a_0$ на булевых значениях переменных. Поэтому в качестве порогового элемента можно взять многочлен $q = q_1 + q_2$, и, без ограничения общности, далее будем рассматривать только такие многочлены для задания квадратичных пороговых функций. Очевидно, что многочлен $q - a_0$ является квадратичной формой. Заметим, что в силу вышеизложенного линейные пороговые булевы функции являются частным случаем квадратичных пороговых функций.

Изменим в целях удобства оперирования квадратичными формами соотношение (3) на равносильное путём переноса свободного члена порогового элемента q в правую часть:

$$f(x_1, \ldots, x_n) = 0 \Leftrightarrow g(x_1, \ldots, x_n) \leqslant t,$$

где $g = q - a_0$; $t = -a_0$.

Укажем способы задания квадратичных пороговых функций. Пусть $G = (g_{ij})$ — матрица коэффициентов квадратичной формы g при фиксированном способе их упорядочивания. Пара (G, t) называется *структурой* квадратичной пороговой функции f (и обозначается $f \sim (G, t)$), элементы матрицы G — весами, а t — порогом. Далее в случаях, допускающих однозначное понимание, для простоты вместо (G, t) будем использовать (g, t). Известно, что структура является неоднозначным способом задания к.п.б.ф. Вместо матрицы коэффициентов многочлена в структуре можно указывать матрицу квадратичной формы. Пусть $g(x_1, \ldots, x_n) = \sum\limits_{1 \leqslant i < j \leqslant n} g_{ij}x_ix_j + \sum\limits_{i=1}^{n} g_ix_i^2$ — каноническая форма многочлена квадратичной формы. Тогда элементы матрицы $W_g = (w_{ij})_{i,j=1,\ldots,n}$ квадратичной формы $g(x_1, \ldots, x_n) = xW_gx^T$ над полем действительных чисел от переменных $x = (x_1, \ldots, x_n)$ определяются как

$$w_{ij} = \frac{g_{ij}}{2}, \quad i < j,$$
$$w_{ij} = w_{ji}, \quad i > j,$$
$$w_{ii} = g_i.$$

Очевидно, что $g_{ij} = w_{ij} + w_{ji}$, $i \neq j$. Квадратичная форма называется канонической (или имеет канонический вид), если её матрица диагональная. Известно, что любая ненулевая квадратичная форма над $\mathbb{R}$ эквивалентна некоторой канонической, т. е. существует такая ортогональная матрица C над $\mathbb{R}$, что замена переменных $x = yC$ переводит матрицу W квадратичной формы в каноническую $K_W = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$, где λ_i — собственные значения матрицы W. Отметим, что при такой замене переменные $y = xC^{-1}$ уже не являются, вообще говоря, булевыми при булевых наборах x, и поэтому эквивалентные квадратичные формы задают в общем случае разные к.п.б.ф. Исключением здесь является случай, в котором булевы наборы переменных x и y связаны невырожденной матрицей перехода над $\mathbb{R}$, тогда эта матрица должна быть подстановочной и соответствующие булевы функции отличаются перестановкой переменных.

Для неподстановочных матриц перехода исходная и новая к.п.б.ф. могут быть не связаны друг с другом линейным преобразованием координат. Аналогично, рассматривая линейное преобразование координат над $\mathrm{GF}(2)$, получаем, что новая функция задается неравенством, возможно, более чем второй степени, в отличие от исходной к.п.б.ф.

Пусть $u = \sum\limits_{i,j=1}^{m} u_{ij}x_ix_j$ — квадратичная форма от m первых переменных, а $v = \sum\limits_{i,j=m+1}^{n} v_{ij}x_ix_j$ — квадратичная форма от $n-m$ остальных переменных. Тогда через $w = u+v$ обозначим естественным образом составленную из этих форм квадратичную форму от n переменных, для которой также будем использовать обозначение через набор весов $w = (u|v)$. Квадратичная форма такого вида называется *распавшейся*.

Лемма 1. Квадратичная форма w является распавшейся тогда и только тогда, когда существует каноническая форма K_w, такая, что матрица перехода C — распавшаяся, т. е. $C = \mathrm{diag}(C_u, C_v)$.

Доказательство. Необходимость следует из того, что матрица W квадратичной формы — распавшаяся, и поэтому каждую клетку можно независимо от другой привести к диагональной с помощью своей матрицы перехода. Достаточность следует из равенства

$$W = \begin{pmatrix} C_u^{-1}K_uC_u & 0 \\ 0 & C_v^{-1}K_vC_v \end{pmatrix}.$$

Лемма доказана. ■

Пусть $A_w = \{\{w(x) : x \in \{0,1\}^n\}\}$ — мультимножество значений квадратичной формы $w(x)$. Через $\{A_w\}$ обозначим множество значений $w(x)$. Под набором $w^* = \left(w_0^*, w_1^* \ldots, w_{2^n-1}^*\right)$ понимается упорядоченный по неубыванию набор элементов мультимножества A_w.

Определим, как в работе [3], понятия нижнего $\lfloor a \rfloor_w$ и верхнего $\lceil a \rceil_w$ приближений действительного числа a в множестве значений $w(x)$ следующим образом:

$$\lfloor a \rfloor_w = \max\{z \in \{A_w\} : z \leqslant a\}; \tag{4}$$

$$\lceil a \rceil_w = \min\{z \in \{A_w\} : z > a\}. \tag{5}$$

Заметим, что нижнее и верхнее приближения a существуют, если и только если $a \geqslant w_0^*$ и $a < w_{2^n-1}^*$ соответственно, где n — ранг w. В дальнейшем для удобства максимальный элемент последовательности w^* обозначим через $w_{\max}^*$. Кроме того, положим $\lfloor a \rfloor_w = -\infty$, если $a < w_0^*$, и $\lceil a \rceil_w = \infty$, если $a \geqslant w_{\max}^*$, и в этом случае будем говорить о бесконечных приближениях.

Очевидны следующие свойства конечных приближений:

1) $\lfloor w_i^* \rfloor_w = w_i^*$, а также $\lfloor w_i^* \rfloor_w = w_j^*$ для всех j, таких, что $w_j^* = w_i^*$;
2) $\lfloor a \rfloor_w \leqslant a$;
3) если $w(x) \leqslant a$, то $w(x) \leqslant \lfloor a \rfloor_w$;
4) $\lceil w_i^* \rceil_w = w_{i+1}^*$, если $w_i^* \neq w_{i+1}^*$;
5) $\lceil a \rceil_w > a$.

Введём в рассмотрение таблицу $S^{(m)} = \|s_{ij}\|$, $i = 0, \ldots, 2^m - 1$, $j = 0, \ldots, 2^{n-m} - 1$, для к.п.б.ф. f со структурой $(u|v,t)$, $\operatorname{rang} u = m$, элементы которой определяются

следующим образом:

$$s_{ij} = 0 \Leftrightarrow u_i^* + v_j^* \leqslant t,$$
$$s_{ij} = 1 \Leftrightarrow u_i^* + v_j^* > t.$$

Для таблицы $S^{(m)}$ справедливо *свойство монотонности*: если $p \geqslant i$, $q \geqslant j$, то $s_{pq} \geqslant s_{ij}$. Легко установить связь таблиц $S^{(m)}$ и $T^{(m)}$. Пусть g_1, g_2 — подстановки степеней 2^m и 2^{n-m}, такие, что $u_i^* = u(g_1(i))$, $v_j^* = v(g_2(j))$, где $i \in \{0, \ldots, 2^m - 1\}$, $j \in \{0, \ldots, 2^{n-m} - 1\}$. Тогда $t_{g_1(i),g_2(j)} = s_{i,j}$.

Неочевидной является связь между функциональной разделимостью с параметром m и диагональным видом матрицы весов к.п.б.ф. Действительно, для $m = 1$ любая, в том числе к.п.б.ф., допускает тривиальную декомпозицию, однако при этом матрица квадратичной формы не обязательно является распавшейся с блоком размера 1. Поэтому в дальнейшем при исследовании функциональной разделимости к.п.б.ф. с параметром m будем рассматривать упрощающее предположение, заключающееся в том, что квадратичная форма — распавшаяся.

2. Критерии функциональной разделимости

Следующая теорема является прямым обобщением теоремы 2 работы [3] и утверждает, что для к.п.б.ф. с распавшейся квадратичной формой при решении вопроса о функциональной разделимости достаточно ограничиться классом квадратичных пороговых булевых функций. Для простоты далее везде будем считать, что перестановка переменных в (1) тождественная.

Теорема 2. К.п.б.ф. f со структурой $(u|v,t)$ допускает простую декомпозицию вида (1) над $F_2(n)$ в том и только в том случае, когда справедливо представление

$$f(x_1, \ldots, x_n) = g(x_1, \ldots, x_m)h_1(x_{m+1}, \ldots, x_n) \vee h_2(x_{m+1}, \ldots, x_n), \tag{6}$$

где g, h_1, h_2 — к.п.б.ф.

Доказательство. Необходимость. Пусть f — к.п.б.ф. со структурой $(u|v,t)$. По теореме 1 в $S^{(m)}$ существует не более двух типов различных строк. Если все строки совпадают, то функция f зависит несущественно от первых m переменных и, полагая $h_1(x_{m+1}, \ldots, x_n) = h_2(x_{m+1}, \ldots, x_n) = f(0, \ldots, 0, x_{m+1}, \ldots, x_n)$, $g(x_1, \ldots, x_m) \equiv 0$, получаем искомое представление (1). Рассмотрим случай, когда в $S^{(m)}$ существуют ровно два типа различных строк. Тогда по свойству монотонности индексы строк одного типа в $S^{(m)}$ образуют последовательность подряд идущих чисел. Пусть строки с индексами $0, 1, \ldots, r-1$ совпадают и более таких строк в $S^{(m)}$ нет. Определим таблицы M_1 и M_2 размера $2^m \times 2^{n-m}$ следующим образом. M_1 имеет нулевыми r первых строк, а остальные совпадают со строкой $\mathbf{s}_r$. M_2 состоит из одинаковых строк, совпадающих со строкой $\mathbf{s}_0$ — первой строкой таблицы $S^{(m)}$. Тогда по свойству монотонности имеем $S^{(m)} = M_1 \vee M_2$, где дизъюнкция (а также ниже используемая операция булева умножения) применяется к таблицам поэлементно. Для M_1 справедливо представление $M_1 = M_1^{(1)} \cdot M_1^{(2)}$, где $M_1^{(1)}$ — таблица размера $2^m \times 2^{n-m}$, состоящая из первых r нулевых и $(2^m - r)$ 1-константных остальных строк; $M_1^{(2)}$ — таблица размера $2^m \times 2^{n-m}$, состоящая из одинаковых строк, совпадающих с $\mathbf{s}_r$. Итак, имеем $S^{(m)} = M_1^{(1)} \cdot M_1^{(2)} \vee M_2$. Напомним, что булевы умножение и дизъюнкция являются линейно-пороговыми функциями, а значит — к.п.б.ф. Поэтому для завершения доказательства достаточно показать, что $M_1^{(1)}$, $M_1^{(2)}$, M_2 определяют табличное задание

искомых к.п.б.ф. g, h_1, h_2 в (6). Пусть $w = (u|v)$ и подстановки g_1 и g_2 задают связь таблиц $S^{(m)}$ и $T^{(m)}$. Приведём подробное изложение доказательства для таблицы M_2. Для $M_1^{(1)}$ и $M_1^{(2)}$ рассуждения практически полностью аналогичны. По построению таблица $M_2 = (m_{ij}^{(2)})$ состоит из одинаковых строк, совпадающих с $\mathbf{s}_0$, следовательно, $T_2 = (t_{ij}^{(2)})$, $t^{(2)}_{g_1(i),g_2(j)} = m^{(2)}_{i,j}$, является таблицей булевой функции, зависящей от переменных $x_{m+1}, \ldots, x_n$. Обозначим эту булеву функцию через h_2. Покажем, что h_2 — к.п.б.ф. Действительно, справедлива следующая цепочка равносильных утверждений:

$$h_2(y) = h_2(y_1, \ldots, y_{n-m}) = 0 \Leftrightarrow \forall\, p \in \{0, \ldots, 2^m - 1\}\ (t^{(2)}_{p,y} = m^{(2)}_{g_1^{-1}(p),g_2^{-1}(y)} = 0) \Leftrightarrow$$
$$\Leftrightarrow s_{0,q} = 0,\ q = g_2^{-1}(y) \Leftrightarrow u_0^* + v_q^* \leqslant t \Leftrightarrow u_0^* + v(g_2(q)) \leqslant t \Leftrightarrow v(y) \leqslant t - u_0^*.$$

Таким образом, h_2 — к.п.б.ф. со структурой $(v, t - u_0^*)$. Аналогично, таблицы $M_1^{(1)}$ и $M_1^{(2)}$ задают к.п.б.ф. g и h_1 со структурами (u, u_{r-1}^*) и $(u, t - u_r^*)$ соответственно.

Достаточность очевидна. Действительно, искомая декомпозиция имеет вид

$$\varphi(x_1, \ldots, x_m) = g, \psi(\varphi, x_{m+1}, \ldots, x_n) = \varphi h_1(x_{m+1}, \ldots, x_n) \vee h_2(x_{m+1}, \ldots, x_n).$$

Теорема доказана. ■

Следующие результаты обобщают теорему 3 работы [3] и сводят исследование функциональной разделимости к.п.б.ф. к анализу её структуры.

Теорема 3. К.п.б.ф. f со структурой $(u|v, t)$ допускает простую декомпозицию вида (1) над $F_2(n)$ в том и только в том случае, когда найдутся такие числа c_1 и c_2, что выполняются неравенства

$$\max(\lfloor c_1 \rfloor_u + \lfloor t - u_0^* \rfloor_v, u_{\max}^* + \lfloor c_2 \rfloor_v) \leqslant t < \lceil c_1 \rceil_u + \lceil c_2 \rceil_v. \tag{7}$$

Доказательство. Необходимость. По теореме 2 справедливо представление (6). Тогда неравенства (7) выполняются для соответствующих приближений порогов $c_1 = u_{r-1}^*$, $c_2 = t - u_r^*$, $c_3 = t - u_0^*$ функций g, h_1, h_2, участвующих в представлении (6). Действительно, неравенство $\lfloor c_1 \rfloor_u + \lfloor c_3 \rfloor_v \leqslant t$ является следствием совпадения строк таблицы $S^{(m)}$ с номерами из множества $\{0, 1, \ldots, r-1\}$, а выполнение неравенства $u_{\max}^* + \lfloor c_2 \rfloor_v \leqslant t < \lceil c_1 \rceil_u + \lceil c_2 \rceil_v$ равносильно совпадению остальных строк $S^{(m)}$.

Достаточность. Если все строки таблицы $S^{(m)}$ для функции f одинаковы, имеем тривиальную декомпозицию для функции f, т. е. то, что и требовалось доказать. Поэтому будем считать, что имеем не менее двух типов различных строк в таблице $S^{(m)}$.

Пусть для заданного m-разбиения $(u|v, t)$ и числа t система (7) имеет решение c_1, c_2. Покажем, что функция f имеет вид (6), причём g, h_1, h_2 — к.п.б.ф. со структурами (u, c_1), (v, c_2), (v, c_3) соответственно, где $c_3 = t - u_0^*$. Правая часть (6) определяет некоторую функцию $p(x_1, \ldots, x_n) = gh_1 \vee h_2$. Докажем, что функции $p(x_1, \ldots, x_n)$ и $f(x_1, \ldots, x_n)$ равны.

Рассмотрим случай $c_2 \geqslant t - u_0^*$. Пусть α — такой индекс столбца, что выполняется $v_\alpha^* \leqslant c_2 < v_{\alpha+1}^*$ (ниже рассмотрим остальные варианты для α). Тогда из (7) имеем $u_{\max}^* + \lfloor c_2 \rfloor_v = u_{\max}^* + v_\alpha^* \leqslant t$, и по свойству монотонности $S^{(m)}$ элементы $s_{i,j} = 0$ для всех i и $j \leqslant \alpha$. С другой стороны, $t \leqslant u_0^* + c_2 < u_0^* + v_{\alpha+1}^*$, что означает $s_{i,j} = 1$ для всех i и $j > \alpha$. Таким образом, в таблице $S^{(m)}$ все строки одинаковые, что противоречит нашим предположениям. Если же $\alpha + 1 = 0$, то $s_{0,0} = 1$ и $f \equiv 1$, что также приводит к противоречию. К аналогичному результату, а именно $f \equiv 0$, приводит предположение $v_\alpha^* = v_{\max}^*$.

Пусть $c_2 < t - u_0^* = c_3$. Функция $p(x_1, \ldots, x_n) = 0$ в том и только в том случае, если для некоторых $y \in \{0,1\}^m$, $z \in \{0,1\}^{n-m}$, $x = (y,z)$ выполняется система

$$\begin{cases} g(y)h_1(z) = 0, \\ h_2(z) = 0, \end{cases} \Leftrightarrow \left(\begin{cases} g(y) = 0, \\ h_2(z) = 0, \end{cases} \text{ или } \begin{cases} h_1(z) = 0, \\ h_2(z) = 0 \end{cases} \right).$$

Тогда, продолжая равносильные преобразования, получим

$$\begin{cases} u(y) \leqslant \lfloor c_1 \rfloor_u, \\ v(z) \leqslant \lfloor c_3 \rfloor_v, \end{cases} \text{ или } \; v(z) \leqslant \lfloor c_2 \rfloor_v,$$

что означает с учётом (7) выполнение в первом случае неравенства

$$w(y,z) = u(y) + v(z) \leqslant \lfloor c_1 \rfloor_u + \lfloor c_3 \rfloor_v \leqslant t,$$

а во втором — $w(y,z) = u(y) + v(z) \leqslant u^*_{\max} + v(z) \leqslant u^*_{\max} + \lfloor c_2 \rfloor_v \leqslant t$.

Итак, доказано включение

$$A_p^0 = \{x \in \{0,1\}^n : p(x) = 0\} \subseteq A_f^0 = \{x \in \{0,1\}^n : f(x) = 0\}. \tag{8}$$

В случае, если $p(y,z) = 1$, имеем

$$(g(y)h_1(z) = 1 \text{ или } h_2(z) = 1) \Leftrightarrow \left(\begin{cases} g(y) = 1, \\ h_1(z) = 1, \end{cases} \text{ или } h_2(z) = 1 \right) \Leftrightarrow$$

$$\Leftrightarrow \left(\begin{cases} u(y) \geqslant \lceil c_1 \rceil_u, \\ v(z) \geqslant \lceil c_2 \rceil_v, \end{cases} \text{ или } v(z) > c_3 \right).$$

Проанализируем последнюю систему. Если выполняется $g(y)h_1(z) = 1$, то по условию получаем

$$w(y,z) = u(y) + v(z) \geqslant \lceil c_1 \rceil_u + \lceil c_2 \rceil_v > t.$$

В случае $v(z) > c_3$ имеем следующую цепочку неравенств:

$$w(y,z) = u(y) + v(z) > u(y) + c_3 = u(y) - u_0^* + t \geqslant t.$$

Итак, из $p(y,z) = 1$ получаем $w(y,z) > t$, что доказывает включение

$$A_p^1 \subseteq A_f^1. \tag{9}$$

Объединяя (8) и (9), учитывая, что $|A_p^0 \cup A_p^1| = |A_f^0 \cup A_f^1| = 2^n$, $A_p^0 \cap A_p^1 = A_w^0 \cap A_w^1 = \varnothing$, получим $A_p^0 = A_f^0$ и $A_p^1 = A_f^1$, то есть f имеет вид (6), что и требовалось доказать. Теорема доказана. ■

Замечание 1. Легко видеть, что теоремы 2 и 3 по существу используют только факт декомпозиции полинома $w(x) = w(y,z) = u(y) + v(z)$. По этой причине эти результаты можно обобщить на произвольный функциональный класс, содержащий дизъюнкцию и конъюнкцию.

Пример 1. Построим с использованием теоремы 3 недекомпозируемую квадратичную пороговую булеву функцию от 10 переменных, задаваемую распавшейся матрицей квадратичной формы. Пусть $\mathbf{1}_n$ — целочисленная квадратная матрица размера n, состоящая из одних единиц. Рассмотрим к.п.б.ф. $f(x_1, \ldots, x_{10})$, задаваемую структурой $(\mathbf{1}_5|\mathbf{1}_5, 30)$. Тогда конечное приближение $\lfloor c \rfloor_{\mathbf{1}_5} = a^2$ для некоторого $a \in \{0, \ldots, 5\}$, $\lceil c \rceil_{\mathbf{1}_5} = (a+1)^2$ и система (7), принимающая вид

$$\begin{cases} a^2 \leqslant 5, \\ b^2 \leqslant 5, \\ (a+1)^2 + (b+1)^2 > 30, \end{cases}$$

очевидно, не имеет решений относительно a и b.

Теорема 4. К.п.б.ф. f со структурой $(u|v,t)$ допускает представление вида (1) над $F_2(n)$ тогда и только тогда, когда

$$\left| \left\{ \lfloor t-b \rfloor_u : b \in \{A_v\},\ u_0^* \leqslant t-b < u_{\max}^* \right\} \right| \leqslant 1. \tag{10}$$

Доказательство. Пусть число переменных функции равно n, а параметр разбиения m. Воспользуемся табличным представлением функции f в виде $S^{(m)}$. В силу условия монотонности индексы одинаковых строк таблицы $S^{(m)}$ образуют последовательность подряд идущих целых чисел. По теореме 1 достаточно доказать, что выполнение неравенства (10) означает существование в $S^{(m)}$ не более двух различных типов строк, и наоборот. Строки с индексами i и $i+1$ различны в том и только в том случае, если найдётся такой $j \in \{0, \ldots, 2^{n-m}-1\}$, что $s_{i,j} < s_{i+1,j}$. Это означает $u_i^* + v_j^* \leqslant t < u_{i+1}^* + v_j^*$, что равносильно существованию $b \in \{A_v\}$, такому, что $u_i^* \leqslant t-b < u_{i+1}^*$. При этом $u_i^* = \lfloor t-b \rfloor_u$. Таким образом, знакоперемена $(0,1)$ в b-м столбце однозначно определяется величиной $\lfloor t-b \rfloor_u < u_{\max}^*$. Следовательно, различные конечные нижние приближения $\lfloor t-b \rfloor_u < u_{\max}^*$ соответствуют различным строкам (при этом первая строка не входит в подсчёт, так как не образует знакоперемены). Таким образом, число таких конечных нижних приближений равно числу блоков одинаковых строк таблицы $S^{(m)}$ минус единица. ■

ЛИТЕРАТУРА

1. *Черемушкин А. В.* Бесповторная декомпозиция сильно зависимых функций // Дискретная математика. 2004. Т. 16. Вып. 3. С. 3–42.
2. *Ежов А. А., Шумский С. А.* Нейрокомпьютинг и его применения в экономике и бизнесе. М.: МИФИ, 1998. 222 с.
3. *Шурупов А. Н.* О функциональной разделимости булевых пороговых функций // Дискретная математика. 1997. Т. 9. Вып. 2. С. 59–73.
4. *Ashenhurst R. L.* The decomposition of switching functions // Ann. Comput. Laborat. Harv. Univ. 1959. V. 29. P. 74–116.
5. *Дертоузос М.* Пороговая логика. М.: Мир, 1967. 343 с.
6. *Логачев О. А., Сальников А. А., Ященко В. В.* Булевы функции в теории кодирования и криптологии. М.: МЦНМО, 2004. 470 с.
7. *Подольский В. В.* Оценки весов персептронов (полиномиальных пороговых булевых функций): автореф. дис. ... канд. физ.-мат. наук. М.: МГУ им. М. В. Ломоносова, 2009.
8. *Crama Y. and Hammer P.* Boolean Functions. Theory, Algorithms and Applications. Cambridge University Press, 2011.

REFERENCES

1. *Cheremushkin A. V.* Iteration-free decomposition of strongly dependent functions. Discr. Math. Appl., 2004, vol. 14, iss. 5, pp. 439–478.
2. *Ezhov A. A., Shumskiy S. A.* Neyrokomp'yuting i ego primeneniya v ekonomike i biznese [Neurocomputing and its Applications in Economics and Business]. Moscow, MEPhI Publ., 1998. 222 p. (in Russian)
3. *Shurupov A. N.* On the functional decomposability of Boolean threshold functions. Discr. Math. Appl., 1997, vol. 7, iss. 3, pp. 257–272.
4. *Ashenhurst R. L.* The decomposition of switching functions. Ann. Comput. Laborat. Harv. Univ., 1959, vol. 29, pp. 74–116.
5. *Dertouzos M.* Threshold Logic: A Synthesis Approach. Cambridge, MA, The M.I.T. Press, 1965.
6. *Logachev O. A., Sal'nikov A. A., Yashchenko V. V.* Bulevy funktsii v teorii kodirovaniya i kriptologii [Boolean Functions in Coding Theory and Cryptology]. Moscow, MCCME Publ., 2004. 470 p. (in Russian)
7. *Podol'skiy V. V.* Otsenki vesov perseptronov (polinomial'nykh porogovykh bulevykh funktsiy) [Ratings Scales Perceptrons (Polynomial Threshold Boolean Functions)]. Abstract of PhD in Physics and Mathematics thesis. Moscow, MSU Publ., 2009. (in Russian)
8. *Crama Y. and Hammer P.* Boolean Functions. Theory, Algorithms and Applications. Cambridge University Press, 2011.

МАТЕМАТИЧЕСКИЕ МЕТОДЫ КРИПТОГРАФИИ

УДК 512.54

КРИПТОГРАФИЧЕСКИЙ АНАЛИЗ ПРОТОКОЛА АУТЕНТИФИКАЦИИ УШАКОВА — ШПИЛЬРАЙНА, ОСНОВАННОГО НА ПРОБЛЕМЕ БИНАРНО СКРУЧЕННОЙ СОПРЯЖЁННОСТИ[1]

М. Н. Горнова, Е. Г. Кукина, В. А. Романьков

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

Приводится криптографический анализ протокола аутентификации Ушакова — Шпильрайна, базирующегося на проблеме бинарно скрученной сопряжённости относительно пары эндоморфизмов полугруппы 2×2-матриц над кольцом срезанных многочленов с коэффициентами из поля $\mathbb{F}_2$. Показано, что закрытый ключ протокола может быть вычислен путём решения системы линейных уравнений над полем $\mathbb{F}_2$. Представлена теоретическая оценка сложности алгоритма данного криптографического анализа и дано краткое описание практических результатов, полученных с помощью подготовленной программы. Анализ показывает, что рассматриваемый протокол аутентификации является теоретически и практически нестойким.

Ключевые слова: *криптография, аутентификация, эндоморфизм, скрученная сопряжённость, срезанные многочлены.*

DOI 10.17223/20710410/28/5

CRYPTANALYSIS OF USHAKOV — SHPILRAIN'S AUTHENTICATION PROTOCOL BASED ON THE TWISTED CONJUGACY PROBLEM

M. N. Gornova, E. G. Kukina, V. A. Romankov

Omsk State University, Omsk, Russia

E-mail: romankov48@mail.ru

We give a cryptanalysis of Ushakov — Shpilrain's authentication protocol based on the twisted conjugacy problem for a pair of endomorphisms on the semigroup of all 2×2 matrices over the ring of truncated one-variable polynomials over the field $\mathbb{F}_2$. It is shown that the private key of the protocol can be computed by solving the system of linear equations over $\mathbb{F}_2$. We present a theoretical estimation for the complexity of this cryptanalysis and describe practical results obtained in a computer experiment. It is shown that the protocol is theoretically and practically vulnerable.

Keywords: *cryptography, authentication, endomorphism, twisted conjugacy, truncated polynomials.*

[1]Работа выполнена при финансовой поддержке РФФИ, проекты № 13.01.00239-а и Р. Сибирь 15-41-04312-а.

Введение

В современной алгебраической криптографии примитивы, схемы, протоколы и системы строятся на алгебраических структурах (платформах). Наиболее развита в этом смысле криптография на бесконечных группах (см. работы [1, 2]). Предположения секретности при этом, как правило, базируются на трудноразрешимых и неразрешимых алгоритмических проблемах. Среди последних чаще всего фигурирует проблема сопряжённости, но встречаются также схемы, использующие проблемы равенства, вхождения и т. п. Анализ таких схем ведётся с точки зрения теории сложности. При этом кроме классического понятия «сложности в худшем случае» используются понятия «сложности в среднем» и «генерической сложности» [1 – 4].

В настоящей работе рассматривается протокол аутентификации из [5], для которого в качестве платформы предлагается использовать полугруппу 2×2-матриц над срезанными многочленами от одной переменной над полем $\mathbb{F}_2$, состоящим из двух элементов. В качестве базовой трудноразрешимой проблемы фигурирует проблема бинарно скрученной сопряжённости. Перейдём к определениям.

Пусть G — (полу)группа, ψ — её эндоморфизм. Говорят, что элементы $u, v \in G$ *скрученно сопряжены* относительно ψ, или, более кратко, ψ-*сопряжены*, если существует элемент $s \in G$, такой, что выполняется равенство

$$\psi(s)u = vs.$$

Легко проверить, что свойство *быть* ψ-*сопряжёнными* является отношением эквивалентности на основном множестве G. Понятие скрученной сопряжённости в случае, когда G — группа, обобщает понятие сопряжённости, соответствующее тождественному эндоморфизму $\psi = id$. Его появление в начале XX столетия мотивировано топологической теорией фиксированных точек отображений Нильсена — Райдемайстера. Впоследствии это понятие нашло применение в различных областях математики: теории представлений бесконечных групп, теории динамических систем, алгебраической геометрии и т. п. Свойство скрученной сопряжённости интересно и с чисто алгебраической точки зрения (см. по этому поводу работы [6 – 11]). Имеются попытки использовать понятие скрученной сопряжённости для алгебраических приложений. Рассмотрим одну из таких попыток [5], в которой фигурирует даже более общее понятие бинарно скрученной сопряжённости.

Пусть G — (полу)группа, ψ, φ — её эндоморфизмы. Говорят, что элементы $u, v \in G$ *скрученно сопряжены* относительно ψ, φ, или, более кратко, ψ-,φ-*сопряжены*, если существует элемент $s \in G$, такой, что выполняется равенство

$$\psi(s)u = v\varphi(s).$$

Свойство *быть* ψ-,φ-*сопряжёнными* также является отношением эквивалентности на основном множестве G. Понятие бинарно скрученной сопряжённости обобщает понятие скрученной сопряжённости, в котором энодоморфизм ψ произвольный, а φ — тождественный.

Говорят, что в G разрешима *проблема скрученной сопряжённости*, если существует алгоритм, определяющий для любого эндоморфизма ψ и любой пары элементов u, v из G, является ли элемент u ψ-сопряжённым элементу v. Аналогично определяется *проблема бинарно скрученной сопряжённости*. В [8] установлена разрешимость проблемы скрученной сопряжённости в произвольной полициклической группе P относительно любого её эндоморфизма ψ. В [9] доказана разрешимость проблемы скрученной сопряжённости в произвольной конечно порождённой метабелевой группе M

относительно любого её эндоморфизма ψ, тождественного по модулю коммутанта M' группы M. Конечно, существуют группы, в которых указанные проблемы алгоритмически неразрешимы. Например, такова конечно определённая группа с неразрешимой классической проблемой сопряжённости. Относительно примеров таких групп см., например, [12, 13].

В криптографии на бесконечных группах предположения секретности обычно связывают с *поисковой* алгоритмической проблемой. Например, известно, что два элемента u, v сопряжены в группе G. Требуется найти (хотя бы один) сопрягающий элемент $f \in G$, такой, что $fuf^{-1} = v$. Если проблема сопряжённости в группе G алгоритмически неразрешима, то время решения соответствующей проблемы поисковой сопряжённости не может быть ограничено сверху никакой рекурсивной функцией от параметров проблемы. Например, если на группе заданы длины элементов, то мы не можем ограничить сверху минимальную длину сопрягающего элемента f рекурсивной функцией от длин элементов u и v. Обычно в качестве длины выступает длина кратчайшего группового слова от фиксированного конечного множества порождающих элементов группы G, представляющего данный элемент. Считается, что трудноразрешимым алгоритмическим проблемам для данной группы соответствуют трудноразрешимые поисковые алгоритмические проблемы. Нахождение и интерпретация таких проблем и соответствующих групп является одной из основных проблем криптографии на бесконечных группах. В этой связи следует заметить, что в [4, 14] приведены многочисленные примеры раскрытия передаваемых секретных сообщений без вычисления закрытых ключей шифрования, т. е. без решения поисковых алгоритмических проблем, на трудности решения которых базируются предположения секретности. Это заставляет пересмотреть общее мнение о построении систем такого вида. Дальнейший криптографический анализ, основанный на разработанном в [4] методе линейного разложения, можно найти в [15 – 17].

В [5] в качестве платформы для построения протокола аутентификации предлагается использовать полугруппу 2×2-матриц над кольцом K_n n-срезанных многочленов от одной переменной x с коэффициентами из поля $\mathbb{F}_2$. Здесь n — натуральное число. Более точно, $K_n = \mathbb{F}_2[x]/I_n$ означает фактор-кольцо кольца многочленов $\mathbb{F}_2[x]$ по идеалу $I_n = \mathrm{ideal}(x^n)$, порождённому элементом x^n. Произвольный элемент $f(x)$ кольца K_n однозначно записывается в виде $a_0 + a_1x + \ldots + a_{n-1}x^{n-1}$, где $a_i \in \mathbb{F}_2$. При этом a_0 называется *свободным членом* данного многочлена. Здесь и в дальнейшем мы опускаем слово n-срезанный. Сложение таких нормальных форм обычное, умножение отличается от обычного тем, что все степени x^k при $k \geqslant n$ равны нулю в K_n, поэтому они не фигурируют в нормальной форме.

1. Протокол аутентификации Ушакова — Шпильрайна

Приведём описание протокола аутентификации Ушакова — Шпильрайна из [5]. Заметим, что в указанной работе приведена общая версия протокола, в которой фигурирует антиэндоморфизм полугруппы. Один из основных частных случаев связан с выбором в качестве такого антиэндоморфизма операции взятия обратного элемента; мы рассматриваем и анализируем именно эту версию протокола.

Подобно классической схеме Фиата — Шамира, предполагается k-кратное повторение раундов при одной сессии аутентификации. В каждом из этих раундов вероятность правильного прохождения при незнании секретного ключа равна $1/2$. Значит, вероятность прохождения при незнании секретного ключа в k последовательных раундах не больше $(1/2)^k$.

Перейдём к описанию протокола. Предположим, что один из корреспондентов (Алиса) доказывает своё право на аутентификацию, а другой (Боб) осуществляет проверку её права на аутентификацию. При этом они открыто договариваются о выборе полугруппы G в качестве платформы для протокола аутентификации и двух эндоморфизмов ψ и φ полугруппы G, участвующих в установке этого протокола.

В процессе установки Алиса выбирает в качестве закрытого ключа обратимый элемент s полугруппы G. Затем она выбирает элемент $w \in G$ и вычисляет $t = \psi(s^{-1})w\varphi(s)$. Открытым ключом Алисы служит пара (t, w).

Каждый раунд аутентификации описывается следующим образом:

1) При очередной сессии аутентификации Алиса выбирает сессионный закрытый ключ r, с помощью которого она вычисляет $u = \psi(r^{-1})t\varphi(r)$, после чего посылает u Бобу.
2) Боб выбирает с вероятностью $1/2$ один из битов $b = 0$ или $b = 1$ и посылает его Алисе.
3) Если Алиса получает $b = 0$, она должна отправить Бобу сессионный ключ $v = r$. При его получении Боб проверяет выполнимость равенства $u = \psi(v^{-1})t\varphi(v)$. Условием прохождения раунда является справедливость этого равенства.
   Если Алиса получает $b = 1$, она должна вычислить и отправить Бобу элемент $v = sr$. При его получении Боб проверяет выполнимость равенства $u = \psi(v^{-1})w\varphi(v)$. Условием прохождения раунда является справедливость этого равенства.

Непосредственно проверяется, что при правильном элементе v указанные равенства действительно справедливы. В то же время любой, кто захочет выдать себя за Алису, сможет это сделать в случае, если угадает значение ответного бита Боба.

Действительно, если он угадает ответ $b = 0$, то ему достаточно выбрать произвольный ключ r, а затем вычислить и передать элемент $u = \psi(r^{-1})t\varphi(r)$. При угаданном ответе $b = 0$ передаётся $v = r$. Но если ответ $b = 1$, необходимо передать элемент $v = rs$, что равносильно раскрытию закрытого ключа s.

Если обманщик угадает ответ $b = 1$, он также может успешно завершить данный раунд процесса аутентификации. Для этого он должен выбрать любой элемент $z \in G$, вычислить и передать Бобу элемент $u = \psi(z^{-1})w\varphi(z)$. При ответе $b = 1$ он просто отправляет Бобу элемент $v = z$ и таким образом проходит проверку. Но при ответе $b = 0$ ему для прохождения проверки нужно передать такой элемент v, что выполнено равенство $u = \psi(v^{-1})t\varphi(v)$. Подходит элемент $v = s^{-1}z$, знание которого позволяет раскрыть s.

Криптостойкость приведённого протокола обеспечивается трудноразрешимостью проблемы бинарно скрученной сопряжённости в выбранной полугруппе G. Действительно, если кому-либо удалось бы найти по элементам t и w закрытый ключ s, то таким образом был бы раскрыт весь протокол. Если по элементам u и t удалось бы вычислить сессионный ключ r, то при ответе $b = 1$ также удалось бы вычислить закрытый ключ s.

Заметим, что достаточно было бы вычислить такой элемент s', для которого выполнено равенство $t = \psi((s')^{-1})w\varphi(s')$. С помощью s' так же можно проходить аутентификацию, как и с оригинальным закрытым ключом s. Если вычислить элемент $r' \in G$, такой, что $\psi((r')^{-1})t\varphi(r') = u$, то можно, зная sr, определить элемент $s' = (sr)(r')^{-1}$.

Тогда

$$\psi((s')^{-1})w\varphi(s') = \psi(r'r^{-1})(\psi(s^{-1})w\varphi(s))\varphi(r(r')^{-1}) =$$
$$= \psi(r')(\psi(r^{-1})t\varphi(r))\varphi((r')^{-1}) = \psi(r')u\varphi((r')^{-1}) = t.$$

Таким образом, s' также играет роль закрытого ключа s. Вычислить s' можно, наблюдая передаваемый Алисой элемент $v = sr$ при ответе Боба $b = 1$.

В [5] в качестве платформы протокола предлагается использовать полугруппу G всех 2×2-матриц над кольцом срезанных многочленов K_n, где n — число порядка 300. Любое отображение $\phi : K_n \to K_n$, для которого $\phi(x) = h$, где $h = h(x)$ — многочлен с нулевым свободным членом, однозначно продолжается до эндоморфизма кольца K_n, рассматриваемого как алгебра над $\mathbb{F}_2$. Обозначим такой эндоморфизм через ϕ_h. Действительно, $\mathbb{F}_2[x]$ — свободная коммутативная ассоциативная алгебра над $\mathbb{F}_2$ размерности один со свободным порождающим x. Поэтому любое отображение $x \mapsto h$, где h — произвольный многочлен из $\mathbb{F}_2[x]$, однозначно продолжается до её эндоморфизма. Если h — многочлен с нулевым свободным членом, то идеал I_n инвариантен относительно этого эндоморфизма. В этом случае данный эндоморфизм индуцирует эндоморфизм ϕ_h кольца K_n. Любой эндоморфизм кольца K_n естественным образом распространяется на полугруппу G, действуя соответствующим образом на элементы матриц. В [5] в качестве фигурирующих в протоколе аутентификации φ и ψ предлагается выбирать соответствующие распространения на G эндоморфизмов кольца K_n вида ϕ_h, где $h \in K_n$ имеет нулевой свободный член. Для этих распространений сохраняются обозначения ϕ_h.

Параметр n определяет размер ключевого пространства для закрытых ключей. При $n = 300$ имеется 2^{300} многочленов степени $\leqslant 300$ над $\mathbb{F}_2$, следовательно, полугруппа G содержит 2^{1200} элементов, что также является размером ключевого пространства. В то же время вычисления в кольце K_n эффективны [18]. Сложение $p(x) + q(x)$ элементов $p(x), q(x) \in K_n$ осуществляется за время $\mathrm{O}(n)$, умножение $p(x)q(x)$ — за время $\mathrm{O}(n \log_2 n)$, композиция $p(q(x))$ — за время $\mathrm{O}(n \log_2 n)$. Отюда получаем оценку времени работы одного раунда алгоритма протокола, которая определяется как $\mathrm{O}(n \log_2 n)^{3/2}$.

Открытыми ключами системы служат пары эндоморфизмов кольца K_n указанного вида. Всего таких пар эндоморфизмов при произвольном параметре $2^{2(n-1)}$. Для $n = 300$ их 2^{598}.

2. Криптографический анализ протокола аутентификации Ушакова — Шпильрайна

Рассмотрим уравнение

$$t = \psi(s^{-1})w\varphi(s), \tag{1}$$

в котором неизвестной является матрица $s = (s(ij))$, $s(ij) \in K_n$, $i, j = 1, 2$. Считаем, как это предлагается авторами протокола в [5], что каждый из эндоморфизмов ψ и φ имеет вид ϕ_h для соответствующих многочленов без свободных членов $h \in K_n$, как это объяснено выше. Умножим уравнение (1) слева на $\psi(s)$. В результате получим равносильное уравнение

$$\psi(s)t = w\varphi(s). \tag{2}$$

Запишем элементы матрицы s в нормальной форме с неопределёнными коэффициентами из поля $\mathbb{F}_2$:

$$s(ij) = s(ij)_0 + s(ij)_1 x + \ldots + s(ij)_{n-1} x^{n-1}.$$

Применив к матрице s эндоморфизмы φ и ψ, получим из уравнения (2) равносильную ему систему линейных уравнений от $4n$ переменных $s(ij)_l$ $(i,j = 1,2;$ $l = 0,1,\ldots,n-1)$ над полем $\mathbb{F}_2$. Остаётся найти такое решение данной системы линейных уравнений, для которого соответствующая ему матрица s обратима. Значительно облегчает эту задачу то обстоятельство, что матрица s обратима тогда и только тогда, когда обратима матрица s_1 свободных членов её элементов, являющаяся образом s относительно гомоморфизма специализации, при котором x отображается в ноль. Действительно, имеется всего шесть обратимых 2×2-матриц над $\mathbb{F}_2$. Следовательно, достаточно рассмотреть шесть различных случаев и хотя бы в одном из них найти обратимое решение s. Каждая частная задача при этом имеет $r = 4(n-1)$ переменных.

3. Алгоритм дешифрования протокола аутентификации Ушакова — Шпильрайна

На входе алгоритма имеем два элемента t и w из полугруппы G матриц размера 2×2 над кольцом K_n, а также два эндоморфизма ψ и φ, имеющие вид ϕ_h, предложенный в протоколе Ушакова — Шпильрайна. Их можно считать эндоморфизмами линейного пространства всей алгебры матриц. Известно, что элементы t и w ψ-,φ-сопряжены, но мы не знаем сопрягающий элемент s. Наша задача — найти любой обратимый элемент s, такой, что $\psi(s)t - w\varphi(s) = 0$. Преобразование $s \mapsto \psi(s)t - w\varphi(s)$ является линейным. Все элементы s, переходящие в ноль (ядро преобразования) образуют подпространство, базис которого находится алгоритмом Гаусса.

Ш а г 1. Получение системы линейных уравнений (СЛУ).

Рассмотрим линейное преобразование $\psi(s)t-w\varphi(s)$ в каждом из шести случаев, отвечающих различным обратимым матрицам специализации s_1, о чём говорилось выше. Каждый такой случай соответствует выбору свободных членов элементов матрицы s. Строим матрицу этого линейного преобразования, выбрав естественный базис полугруппы G. Матрице соответствует СЛУ, состоящая из $4(n-1)$ уравнений от $4(n-1)$ неизвестных. Хорошо известно, что число операций над элементами любого поля для решения такой системы оценивается как $\mathrm{O}(4^3(n-1)^3) \sim \mathrm{O}(n^3)$.

Ш а г 2. Нахождение решения s.

Полученные в шести случаях СЛУ решаем методом Гаусса. По условию хотя бы в одном из этих случаев СЛУ имеет решение. Ему соответствует искомая обратимая матрица s.

Ш а г 3. Проверка.

Считаем s^{-1} и затем $t' = \varphi(s^{-1})w\psi(s)$. Сравниваем t' с t.

Как уже отмечалось, число операций над элементами поля для решения каждой из рассматриваемых СЛУ оценивается как $\mathrm{O}(4^3(n-1)^3) \sim \mathrm{O}(n^3)$. Для получения общей оценки сложности следует учесть сложность элементарных операций в поле $\mathbb{F}_2$. Заметим, что возникающие СЛУ имеют специальный вид, при котором неизвестные коэффициенты при больших степенях переменной x в K_n не участвуют (присутствуют с нулевыми коэффициентами) в уравнениях, соответствующих меньшим степеням переменной x. Система, таким образом, имеет ступенчатый вид, значит, соответствующие вычисления упрощаются.

В соответствии с полученным алгоритмом написана программа на языке Java. Программа запускалась на оборудовании с двухъядерным процессором частоты 2,6 ГГц и 8 Гб оперативной памяти.

Для степени срезанных многочленов порядка 300 программа работает в среднем 7 мин 40 с, причём это время можно сократить, если оптимизировать некоторые шаги и запустить её на более мощном оборудовании.

ЛИТЕРАТУРА

1. *Myasnikov A., Shpilrain V., and Ushakov A.* Group-Based Cryptography. Advances courses in Math., CRM, Barselona. Basel, Berlin, New York: Birkhäuser Verlag, 2008. 183 p.
2. *Myasnikov A., Shpilrain V., and Ushakov A.* Non-commutative Cryptography and Complexity of Group-Theoretic Problems. Amer. Math. Soc. Surveys and Monographs. Providence R.I.: Amer. Math. Soc., 2011. 385 p.
3. *Романьков В. А.* Диофантова криптография на бесконечных группах // Прикладная дискретная математика. 2012. № 2(16). С. 15–42.
4. *Романьков В. А.* Алгебраическая криптография. Омск: ОмГУ, 2013. 135 с.
5. *Shpilrain V. and Ushakov A.* An authentication scheme based on the twisted conjugacy problem // ACNS'2008. LNCS. 2008. V. 5037. P. 366–372.
6. *Fel'shtyn A. and Troitsky E.* Twisted Burnside — Frobenius theory for discrete groups // J. Reine Angew. Math. 2007. V. 613. P. 193–210.
7. *Goncalves D. and Wong P.* Twisted conjugacy classes in nilpotent groups // J. Reine Angew. Math. 2009. V. 633. P. 11–27.
8. *Roman'kov V.* The twisted conjugacy problem in polycyclic groups // J. Group Theory. 2010. V. 13. No. 3. P. 353–364.
9. *Вентура Э., Романьков В. А.* Проблема скрученной сопряжённости для эндоморфизмов метабелевых групп // Алгебра и логика. 2009. Т. 48. № 2. С. 157–173.
10. *Roman'kov V.* Twisted conjugacy classes in nilpotent groups // J. Pure Appl. Algebra. 2011. V. 215. No. 4. P. 664–671.
11. *Fel'shtyn A. and Goncalves D. L.* Reidemeister spectrum for metabelian groups // Int. J. Algebra Comput. 2011. V. 21. No. 3. P. 1–16.
12. *Ремесленников В. Н., Романьков В. А.* Теоретико-модельные и алгоритмические проблемы теории групп // Итоги науки и техн. Сер. Алгебра. Геометрия. Топология. Т. 21. М.: ВИНИТИ, 1983. С. 3–79.
13. *Miller C. F.* Decision problems for groups: survey and reflections // Algorithms and Classification in Combinatorial Group Theory /eds. G. Baumslag and C. F. Miller III. MSRI Publications, 1992. V. 23. P. 1–59.
14. *Романьков В. А.* Криптографический анализ некоторых схем шифрования, использующих автоморфизмы // Прикладная дискретная математика. 2013. № 3(21). С. 36–51.
15. *Roman'kov V. and Myasnikov A.* A linear decomposition attack. `arXiv:1412.6401v1[math.GR]`. 19 Dec. 2014.
16. *Roman'kov V.* A polynomial time algorithm for the braid double shielded public key cryptosystem. `arXiv:1412.5277v1[math.GR]`. 17 Dec. 2014.
17. *Roman'kov V.* Linear decomposition attack on public key exchange protocols using semidirect products of (semi)groups. `arXiv:1501.0052v1[cs.CR]`. 15 Jan. 2015.
18. *Bürgisser P., Clausen M., and Shokrollalu M. A.* Algebraic Complexity Theory. Berlin: Springer, 1997.

REFERENCES

1. *Myasnikov A., Shpilrain V., and Ushakov A.* Group-Based Cryptography. Advances courses in Math., CRM, Barselona. Basel, Berlin, New York: Birkhäuser Verlag, 2008. 183 p.

2. *Myasnikov A., Shpilrain V., and Ushakov A.* Non-commutative Cryptography and Complexity of Group-Theoretic Problems. Amer. Math. Soc. Surveys and Monographs. Providence R.I.: Amer. Math. Soc., 2011. 385 p.
3. *Roman'kov V A.* Diofantova kriptografiya na beskonechnykh gruppakh [Diophantine cryptography over infinite groups.] Prikladnaya Diskretnaya Matematika, 2012, no. 2(16), pp. 15–42. (in Russian)
4. *Roman'kov V. A.* Algebraicheskaya kriptografiya [Algebraic Cryptography]. Omsk: OmSU Publ., 2013. 135 p. (in Russian)
5. *Shpilrain V. and Ushakov A.* An authentication scheme based on the twisted conjugacy problem. ACNS'2008. LNCS, 2008, vol. 5037, pp. 366–372.
6. *Fel'shtyn A. and Troitsky E.* Twisted Burnside — Frobenius theory for discrete groups. J. Reine Angew. Math., 2007, vol. 613, pp. 193–210.
7. *Goncalves D. and Wong P.* Twisted conjugacy classes in nilpotent groups. J. Reine Angew. Math., 2009, vol. 633, pp. 11–27.
8. *Roman'kov V.* The twisted conjugacy problem in polycyclic groups. J. Group Theory, 2010, vol. 13, no. 3, pp. 353–364.
9. *Ventura E. and Roman'kov V. A.* The twisted conjugacy problem for endomorphisms of metabelian groups. Algebra and Logic, 2009, vol. 48, iss. 2, pp. 89–98.
10. *Roman'kov V.* Twisted conjugacy classes in nilpotent groups. J. Pure Appl. Algebra, 2011, vol. 215, no. 4, pp. 664–671.
11. *Fel'shtyn A. and Goncalves D. L.* Reidemeister spectrum for metabelian groups. Int. J. Algebra Comput., 2011, vol. 21, no. 3, pp. 1–16.
12. *Remeslennikov V N., Roman'kov V. A.* Teoretiko-model'nye i algoritmicheskie problemy teorii grupp [Model-theoretic and algorithmic problems in group theory]. Itogi Nauki i Tekhn. Ser. Algebra. Geometriya. Topologiya, vol. 21. Moscow, VINITI Publ., 1983. pp. 3–79. (in Russian)
13. *Miller C. F.* Decision problems for groups: survey and reflections. Algorithms and Classification in Combinatorial Group Theory (eds. G. Baumslag and C. F. Miller III). MSRI Publications, 1992, vol. 23, pp. 1–59.
14. *Roman'kov V A.* Kriptograficheskiy analiz nekotorykh skhem shifrovaniya, ispol'zuyushchikh avtomorfizmy [Cryptanalysis of some schemes applying automorphisms]. Prikladnaya Diskretnaya Matematika, 2013, no. 3(21), pp. 36–51. (in Russian)
15. *Roman'kov V. and Myasnikov A.* A linear decomposition attack. `arXiv:1412.6401v1[math.GR]`. 19 Dec. 2014.
16. *Roman'kov V.* A polynomial time algorithm for the braid double shielded public key cryptosystem. `arXiv:1412.5277v1[math.GR]`. 17 Dec. 2014.
17. *Roman'kov V.* Linear decomposition attack on public key exchange protocols using semidirect products of (semi)groups. `arXiv:1501.0052v1[cs.CR]`. 15 Jan. 2015.
18. *Bürgisser P., Clausen M., and Shokrollalu M. A.* Algebraic Complexity Theory. Berlin, Springer, 1997.

УДК 510.52

О ГЕНЕРИЧЕСКОЙ СЛОЖНОСТИ ПРОБЛЕМЫ РАСПОЗНАВАНИЯ КВАДРАТИЧНЫХ ВЫЧЕТОВ[1]

А. Н. Рыбалов

Омский государственный университет им. Ф. М. Достоевского, г. Омск, 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 классической проблемы распознавания квадратичных вычетов в группах вычетов. Доказывается, что её естественная подпроблема генерически трудноразрешима (то есть трудна для почти всех входов) при условии, что проблема распознавания квадратичных вычетов трудноразрешима в классическом смысле.

Ключевые слова: *генерическая сложность, квадратичный вычет, вероятностный алгоритм.*

DOI 10.17223/20710410/28/6

ON GENERIC COMPLEXITY OF THE QUADRATIC RESIDUOSITY PROBLEM

A. N. Rybalov

Omsk State University, Omsk, Russia

E-mail: alexander.rybalov@gmail.com

Generic-case approach to algorithmic problems was suggested by Myasnikov, Kapovich, Schupp and Shpilrain in 2003. This approach studies behavior of an algorithm on typical (almost all) inputs and ignores the rest of inputs. Many classical undecidable or hard algorithmic problems become feasible in the generic case. But there are generically hard problems. For example, this is the classical discrete logarithm problem. In this talk we consider generic complexity of the quadratic residuosity problem. We fit this problem in the frameworks of generic complexity and prove that its natural subproblem is generically hard provided that the quadratic residuosity problem is hard in the worst case.

Keywords: *generic complexity, quadratic residue, probabilistic algorithm.*

Введение

В работе [1] была развита теория генерической сложности вычислений. В рамках этого подхода алгоритмическая проблема рассматривается не на всем множестве входов, а на некотором подмножестве «почти всех» входов. Такие входы образуют так называемое генерическое множество. Понятие «почти все» формализуется введением естественной меры на множестве входных данных. С точки зрения практики алгоритмы, решающие быстро проблему на генерическом множестве, так же хороши, как и

[1]Работа поддержана грантом РФФИ № 15-41-04312.

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. Многие классические NP-полные проблемы в генерическом случае оказываются легкоразрешимыми [4].

С точки зрения современной криптографии интересны такие алгоритмические проблемы, которые, являясь (гипотетически) трудными в классическом смысле, остаются трудными и в генерическом смысле, т. е. для почти всех входов. Это объясняется тем, что при случайной генерации ключей в криптографическом алгоритме происходит генерация входа некоторой трудной алгоритмической проблемы, лежащей в основе алгоритма. Если проблема генерически легкоразрешима, то для почти всех таких входов её можно быстро решить, и ключи почти всегда будут нестойкими. Поэтому проблема должна быть генерически трудной. Например, для проблемы дискретного логарифма такие результаты получены в [5].

В данной работе изучается генерическая сложность классической проблемы распознавания квадратичных вычетов в группах вычетов. Эта алгоритмическая проблема восходит ещё к Гауссу и является хорошо известной в криптографии (см., например, [6]). До сих пор не известно полиномиальных алгоритмов её решения. В данной работе доказывается, что эта проблема неразрешима за полиномиальное время на любых полиномиальных генерических множествах входов при условии отсутствия полиномиальных вероятностных алгоритмов её решения в худшем случае. Более того, существует правдоподобная гипотеза о том, что любой полиномиальный вероятностный алгоритм можно эффективно дерандомизировать, т. е. построить полиномиальный детерминированный алгоритм, решающий ту же задачу. Хотя гипотеза пока не доказана, имеются серьёзные доводы в её пользу [7].

При доказательстве основного результата использованы методы, развитые в [8, 9].

1. Генерические и пренебрежимые множества

Пусть I — некоторое множество входов. На множестве I определена функция размера $\mathrm{size} : I \to \mathbb{N}$, сопоставляющая каждому элементу $a \in I$ его размер $\mathrm{size}(a)$. Допустим, что для любого n множество I_n элементов из I размера n конечно. Для любого подмножества $S \subseteq I$ определим следующую последовательность:

$$\rho_n(S) = \frac{|S \cap I_n|}{|I_n|}, \quad n = 1, 2, 3, \ldots$$

Величина $\rho_n(S)$ — это вероятность получить вход из множества S при случайной и равномерной генерации элементов из I_n. *Асимптотической плотностью* S назовём 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. Понятие генерического множества является некоторой формализацией интуитивного понятия множества «почти всех» элементов множества I в том смысле, что при увеличении размера элемента вероятность попасть в генерическое множество при случайной и равновероятной генерации элементов стремится к 1.

Алгоритмическая проблема распознавания множества $S \subseteq I$ *генерически полиномиально разрешима*, если существует множество $G \subseteq I$, такое, что

1) G — генерическое;
2) G — разрешимое за полиномиальное время;
3) $G \cap S$ — разрешимое за полиномиальное время.

Генерический алгоритм, решающий проблему S, работает на входе $x \in I$ следующим образом. Сначала определяет, принадлежит ли x генерическому множеству G. Если да, то проверяет принадлежность входа S. Если нет, то отвечает НЕ ЗНАЮ. Такой алгоритм правильно решает проблему распознавания S на почти всех входах.

2. Проблема распознавания квадратичных вычетов

Пусть $\mathbb{Z}/(m)$ — мультипликативная группа вычетов по модулю $m \in \mathbb{N}$. Напомним, что квадратичным вычетом в группе $\mathbb{Z}/(m)$ называется любой элемент x, для которого существует $y \in \mathbb{Z}/(m)$, такой, что $x = y^2$. В противном случае элемент x называется квадратичным невычетом. Под проблемой распознавания квадратичных вычетов понимается проблема распознавания следующего множества:

$$QR = \{(m, x) \in \mathbb{N}^2 : m = pq, \text{ где } p, q \text{ — простые числа,}\\ x \text{ — квадратичный вычет в } \mathbb{Z}/(m)\}.$$

В настоящее время неизвестно полиномиальных алгоритмов (в том числе и вероятностных), решающих проблему распознавания квадратичных вычетов для всех таких модулей m. Более того, на предположении о её трудноразрешимости основаны некоторые криптографические алгоритмы [6].

Для изучения генерической сложности этой проблемы необходимо провести некоторую стратификацию на множестве входов. Рассмотрим любую бесконечную последовательность натуральных чисел $\mu = \{m_1, m_2, \ldots\}$, удовлетворяющую следующим условиям:

1) $2^n < m_n < 2^{n+1}$ для любого n;
2) $m_n = pq$, где p, q — различные простые числа, для любого $n > 1$.

Будем называть такую последовательность *экспоненциальной*. Из знаменитого постулата Бертрана, доказанного П. Л. Чебышевым, следует, что экспоненциальные последовательности существуют. Определим алгоритмическую проблему $QR(\mu)$ как ограничение проблемы распознавания квадратичных вычетов QR на следующее множество входных данных:

$$I = \{(m, x) : m \in \mu, x \in \mathbb{Z}/(m)\}.$$

Под размером входа (m, x) понимается количество бит в двоичной записи числа m минус 1. Заметим, что множество I_n входов проблемы $QR(\mu)$ размера n состоит из всех пар (m, x), где m — единственное число $m \in \mu$, удовлетворяющее условию $2^n < m < 2^{n+1}$, а x — любой элемент из $\mathbb{Z}/(m)$.

Таким образом, $QR(\mu)$ является подпроблемой проблемы QR. Тем не менее можно доказать, что среди проблем $QR(\mu)$ существуют проблемы такие же сложные, как и оригинальная проблема QR.

Лемма 1. Если не существует полиномиального вероятностного алгоритма для проблемы QR, то найдётся такая экспоненциальная последовательность μ, что и для проблемы $QR(\mu)$ не существует полиномиального вероятностного алгоритма.

Доказательство. Пусть $P_1, P_2, \ldots$ — все полиномиальные вероятностные алгоритмы. Из предположения о том, что не существует полиномиального вероятностного алгоритма для проблемы QR, следует, что для любого алгоритма P_n существует бесконечно много групп $\mathbb{Z}/(m)$, в которых он не может решить QR. Из этого следует, что можно выбрать последовательность натуральных чисел $\mu' = \{m_1, m_2, \ldots\}$, являющихся произведениями двух простых, так, чтобы алгоритм P_n не решал QR в группе $\mathbb{Z}/(m_n)$ и для любого n выполнялось $m_{n+1} > 2m_n$. Последовательность μ' можно расширить до экспоненциальной последовательности μ, добавив, где нужно, новые члены. Заметим теперь, что $QR(\mu)$ и есть та проблема, для которой не существует полиномиального вероятностного алгоритма. ■

3. Основной результат

Теорема 1. Если проблема $QR(\mu)$ генерически полиномиально разрешима, то существует полиномиальный вероятностный алгоритм, решающий $QR(\mu)$ для всех входов.

Доказательство. Допустим, существует генерический полиномиальный алгоритм $\mathcal{A}$, разрешающий проблему $QR(\mu)$ на некотором полиномиальном генерическом множестве G. Построим вероятностный полиномиальный алгоритм $\mathcal{B}$, решающий $QR(\mu)$ на всём множестве входов. Алгоритм $\mathcal{B}$ на входе (m, x) работает следующим образом:

1) Проверяет, принадлежит ли (m, x) множеству G. Это делается за полиномиальное время, так как множество G разрешимо за полиномиальное время. Если $(m, x) \in G$, то с помощью алгоритма $\mathcal{A}$ определяет принадлежность множеству $QR(\mu)$. Если нет, переходит к шагу 2.
2) Генерирует случайно и равновероятно элемент $y \in \mathbb{Z}/(m)$. Вычисляет $z = xy^2$.
3) Проверяет, принадлежит ли (m, z) множеству G.
4) Если $(m, z) \in G$, то с помощью алгоритма $\mathcal{A}$ определяет, является ли z квадратичным вычетом. Очевидно, что $z = xy^2$ является квадратичным вычетом тогда и только тогда, когда таковым является x.
5) Если $(m, z) \notin G$, выдаёт ответ НЕТ.

Заметим, что полиномиальный вероятностный алгоритм $\mathcal{B}$ может выдать неправильный ответ только на шаге 5. Докажем, что вероятность этого меньше $1/2$. В группе $\mathbb{Z}/(m)$, где $m = pq$ с простыми p и q, ровно $1/4$ элементов являются квадратами, а потому $z = xy^2$ может равновероятно принимать $1/4$ значений в $\mathbb{Z}/(m)$. В то же время с ростом размера входа для более $3/4$ элементов $u \in \mathbb{Z}/(m)$ входы (m, u) попадают в генерическое множество G. Поэтому вероятность шага 5 стремится к нулю с ростом размера входа и становится меньше $1/2$. ■

Непосредственным следствием теоремы 1 является следующее утверждение.

Теорема 2. Если для проблемы QR не существует полиномиального вероятностного алгоритма, то существует экспоненциальная последовательность μ, такая, что проблема $QR(\mu)$ не является генерически полиномиально разрешимой.

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. *Blum M. and Micali S.* How to generate cryptographically strong sequences of pseudorandom bits // SIAM J. Comput. 1984. V. 13. No. 4. P. 850–864.
6. *Mao B.* Современная криптография: теория и практика. М.: Вильямс, 2005. 768 с.
7. *Impagliazzo R. and Wigderson A.* P=BPP unless E has subexponential circuits: derandomizing the XOR Lemma // Proc. 29th STOC. El Paso: ACM, 1997. P. 220–229.
8. *Myasnikov A. and Rybalov A.* Generic complexity of undecidable problems // J. Symbolic Logic. 2008. V. 73. No. 2. P. 656–673.
9. *Rybalov A.* Generic complexity of Presburger Arithmetic // Theory Comput. Systems. 2010. V. 46. No. 1. P. 2–8.

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. *Blum M. and Micali S.* How to generate cryptographically strong sequences of pseudorandom bits. SIAM J. Comput., 1984, vol. 13, no. 4, pp. 850–864.
6. *Mao V.* Sovremennaya kriptografiya: teoriya i praktika [Modern Cryptography: Theory and Practice]. Moscow, Wil'yams Publ., 2005. 768 p. (in Russian)
7. *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.
8. *Myasnikov A. and Rybalov A.* Generic complexity of undecidable problems. J. Symbolic Logic, 2008, vol. 73, no. 2, pp. 656–673.
9. *Rybalov A.* Generic complexity of Presburger Arithmetic. Theory Comput. Systems, 2010, vol. 46, no. 1, pp. 2–8.

МАТЕМАТИЧЕСКИЕ ОСНОВЫ КОМПЬЮТЕРНОЙ БЕЗОПАСНОСТИ

УДК 004.75

РАЗРАБОТКА БЕЗОПАСНОГО ПРОТОКОЛА РАСПРЕДЕЛЁННОЙ СИСТЕМЫ ПРОВЕДЕНИЯ СОРЕВНОВАНИЙ CTF

Н. И. Анисеня

Национальный исследовательский Томский государственный университет, г. Томск, Россия

Работа посвящена разработке математического метода проведения соревнований Capture the Flag, основанных на решении заданий при угрозе DDoS-атак на сервер организаторов. Предлагается распределённый протокол проведения соревнований, который перекладывает часть функций организаторов на участников. Участники соревнования конкурируют друг с другом и не хотят помогать командам-соперникам, поэтому к защищённости протокола предъявляются высокие требования. Получен протокол, удовлетворяющий поставленным требованиям, рассмотрены атаки, исследована устойчивость протокола к ним, предложены модификации. Описаны возможные направления дальнейших исследований в данной области.

Ключевые слова: *распределённые протоколы, защищённые вычисления, отказоустойчивые системы.*

DOI 10.17223/20710410/28/7

DEVELOPING SAFE PROTOCOL FOR DISTRIBUTED TASK-BASED CTF HOLDING SYSTEM

N. I. Anisenya

National Research Tomsk State University, Tomsk, Russia

E-mail: anisenya@gmail.com

The paper is devoted to mathematical method of holding task-based CTF contests under threat of DDoS attack against organizers' servers. For the purpose, a decentralized protocol is proposed where a part of organizer role is distributed among participants. There are some important security requirements to the protocol due to the competitive nature of CTF contests. The result protocol meets these requirements. Its stability against considered attacks is researched. Directions of further research are described.

Keywords: *distributed protocol, secure computation, fault-tolerant system.*

Введение

Соревнования по компьютерной безопасности Capture The Flag (CTF) [1] в последнее время стали очень популярны. Соревнования CTF имеют различные форматы

проведения [2]: задания на соревнованиях формата Jeopardy зачастую требуют детального ознакомления с новыми уязвимостями, а соревнования в формате Attack-Defense, помимо навыков нападения, дают практический опыт по защите компьютерных систем, что делает чрезвычайно полезным участие в подобных соревнованиях студентов, обучающихся по связанным с компьютерной безопасностью специальностям. Соревнования CTF нередко проводятся в рамках конференций, посвященных компьютерной безопасности (DEFCON [3], PHDAYS [4]), зачастую являясь их главным элементом (RuCTF [5]).

Команда по компьютерной безопасности SiBears Томского государственного университета имеет собственный сервис Blackbox [6, 7] для проведения соревнования CTF формата Jeopardy. С использованием Blackbox проведены множество соревнований, среди которых SiBCTF [8] и отборочные RuCTF 2012 Quals проводились на международном уровне.

Постоянно растущий интерес к CTF-движению привёл к тому, что подобными соревнованиями стали интересоваться злоумышленники. В 2012 г. на отборочных соревнованиях RuCTF 2012 Quals имел место инцидент, когда после дисквалификации за нарушения одной из команд на сервер организаторов была осуществлена DDoS-атака. Её результатом стала недоступность сервера соревнования на некоторое время, в течение которого организаторами настраивалась фильтрация соединений по белому списку IP-адресов, присылаемых честными участниками.

Использованный подход не обладает достаточной гибкостью: командам необходимо заранее знать IP-адреса своих участников, что в случае динамических адресов не всегда возможно, а добавление новых адресов в список осуществляется организаторами вручную. В данной работе предлагается способ борьбы с подобными ситуациями, который заключается в отказе от централизованного взаимодействия участников через сервер организаторов. Этот способ подразумевает распределение части роли организаторов между участниками соревнования, что повышает требования к защищённости вычислений. Рассматриваются атаки, предлагаются улучшения протокола, а также рассматриваются возможные направления дальнейших исследований.

1. Цель, задача и актуальность

Целью работы является предложение математического способа обеспечения доступности соревнования CTF, проводимого в формате Jeopardy, при угрозе DDoS-атаки на организаторов.

Под *централизованным сетевым взаимодействием*, или просто *централизованным взаимодействием*, участников будем понимать такое их взаимодействие, которое полагается на некоторый известный всем участникам узел сети — посредника, имеющего отличную от прочих участников и незаменимую по отношению к ним роль.

Злоумышленником назовём нечестного участника соревнования, который преследует хотя бы одну из следующих целей:

1) нарушение работоспособности системы, с высокой вероятностью приводящее к невозможности участия в соревновании всех участников;
2) нарушение работоспособности системы, с высокой вероятностью приводящее к искажению собственных результатов;
3) нарушение работоспособности системы, с высокой вероятностью приводящее к искажению результатов другого конкретного участника.

Активным участием некоторого узла назовём такое его поведение в сети, при котором он отправляет в сеть данные.

Для достижения указанной цели ставится следующая задача: разработать протокол распределённого проведения соревнований CTF, основанных на решении заданий. Требования к протоколу:

1) в результате работы протокола должна формироваться таблица результатов, позволяющая восстановить очерёдность получения ответов;
2) протокол не должен требовать активного участия организаторов во время проведения соревнования;
3) протокол не должен полагаться на централизованное взаимодействие участников во время проведения соревнования;
4) протокол должен позволять проводить соревнование даже при отключении большого количества участников;
5) протокол должен позволять проводить соревнование даже при большом количестве злоумышленников (нечестных участников).

Поскольку данные в распределённой сети распространяются с ощутимой задержкой, определим следующие отношения для значений времени.

Два значения времени t_1 и t_2 равны с учётом временного окна w, если $|t_1 - t_2| \leqslant w$; обозначим $t_1 =_w t_2$.

Значение времени t_1 предшествует значению времени t_2 с учётом временного окна w, если $t_2 - t_1 > w$; обозначим $t_1 <_w t_2$.

Если $t_1 <_w t_2$ либо $t_1 =_w t_2$, то будем записывать этот факт следующий образом: $t_1 \leqslant_w t_2$.

При разработке протокола не рассматривались следующие ситуации и проблемы:

1) проблемы подготовительного этапа (регистрации команд);
2) ситуация распада графа сети на компоненты связности;
3) недостаточная точность временных расчетов с учётом выбранного временного окна w.

Подход с использованием ресурсов пользователей для поддержания работы единой системы давно применяется в различных сферах, где требуются независимые от доверенной стороны («центра») передача данных, их хранение и обработка. Исследования подобных методов с целью применения в конкурентных средах, то есть в том случае, когда участники распределённой системы не заинтересованы в помощи друг другу, видится важной задачей, актуальной для различных областей деятельности.

2. Начальные условия

Ввиду отсутствия единого сервера участники должны иметь возможность самостоятельно проверять правильность полученного ответа на задание. После того как правильный ответ получен, участник должен передать остальным доказательство получения ответа, не раскрывая самого ответа. Помимо доказательства знания ответа, участник должен предоставить подтверждение того, что ответ получен в определённое время.

Сеть соревнования представляет собой оверлейную p2p-сеть (рис. 1), в которой каждая команда-участник (в том числе и команда организаторов) соединена с некоторыми другими командами-участниками (соседями) напрямую, например посредством TCP-соединения.

Предполагается, что каждый участник сети имеет список идентификаторов команд — открытых ключей. Соответствие идентификаторов командам знают только организаторы.

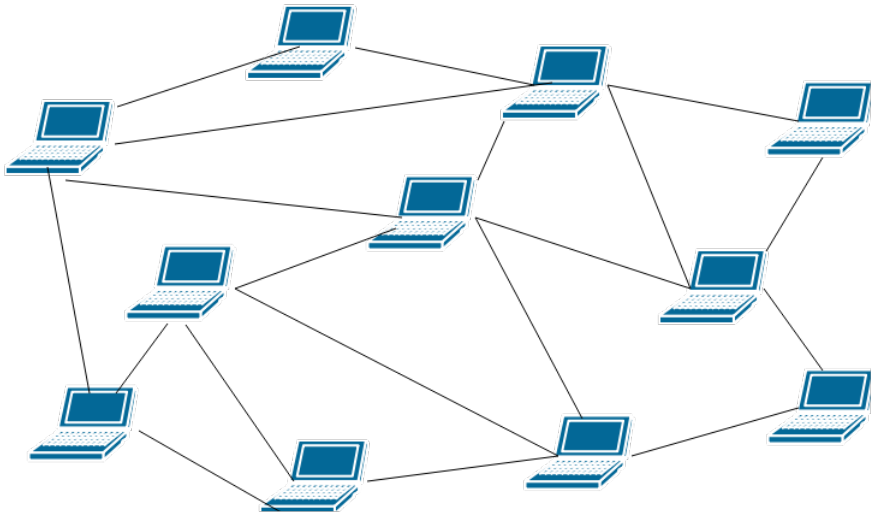


Рис. 1. Иллюстрация p2p-сети

Передача сообщения от Алисы к Бобу в общем случае осуществляется через промежуточные узлы, назовём их *посредниками*. Для отправки сообщения Бобу Алиса сначала шифрует его на открытом ключе Боба. В шифруемое сообщение включается преамбула, позволяющая определить, что процесс расшифрования прошел успешно. Затем Алиса посылает сообщение в сеть, отправляя его каждому своему соседу. Каждый сосед после получения сообщения от Алисы пытается его расшифровать. Очевидно, что успешно расшифровать сообщение способен только Боб, поэтому в случае неуспешного расшифрования текущий узел передаёт полученное сообщение всем своим соседям. Так происходит до тех пор, пока сообщение не будет доставлено Бобу. По аналогии с Freenet [9], для предотвращения бесконечной передачи сообщений узлы не передают одни и те же сообщения дважды.

Описанный способ организации сети соревнования обеспечивает анонимность участников, которая, в свою очередь, с одной стороны, не позволяет участникам блокировать передачу сообщений от или для известной команды, с другой — делает неотличимым от остальных узлов участие в сети команды организаторов.

3. Маршрутизация в условиях конкуренции

3.1. Проблема отказа от посредничества

Участники сети — команды — конкурируют друг с другом и не заинтересованы в доставке чужих сообщений. Для доставки сообщения Бобу Алисе потребуется передать его через посредников. Для защиты конфиденциальности и целостности сообщения оно шифруется на известном Алисе открытом ключе Боба. Однако нет гарантий, что сообщение будет доставлено посредниками. Далее предлагается способ сделать невозможным отказ от посредничества участника сети. Иными словами, если участник сети отказывается поддерживать её работу — выступать в роли посредника, то он не может получить к ней доступ.

3.2. Луковая маршрутизация в сети Tor

Сеть Tor [10] разрабатывалась с целью анонимного доступа к веб-ресурсам. Клиентское приложение Tor представляет собой маршрутизатор, работающий как прокси-сервер. После подключения к сети маршрутизатор получает список IP-адресов узлов сети Tor. Для отправки запроса на веб-ресурс маршрутизатор случайным образом

выбирает нескольких (по умолчанию трёх) посредников из списка узлов, выстраивая таким образом цепочку прокси-серверов. Сообщение шифруется на ключах узлов, состоящих в цепочке, в порядке, обратном порядку следования сообщения: для отправки сообщения m от Алисы к Бобу через посредников M_1, M_2, M_3 оно будет зашифровано следующим образом:

$$E_{M_1}(E_{M_2}(E_{M_3}(m, ip_{Bob}), ip_{M_3}), ip_{M_2}), \tag{1}$$

где ip_x — это IP-адрес узла x. Полученную «луковицу» Алиса отправляет первому посреднику M_1. Получив сообщение (1), посредник M_1 расшифровывает его, как бы снимая свой слой луковицы, и передаёт следующему посреднику M_2 сообщение вида

$$E_{M_2}(E_{M_3}(m, ip_{Bob}), ip_{M_3}). \tag{2}$$

Так продолжается до тех пор, пока Боб не получит исходное сообщение m. Ответ от Боба проходит по той же цепочке в обратном направлении. Разработчиками Tor описанному способу маршрутизации дано название *луковая маршрутизация* (рис. 2).

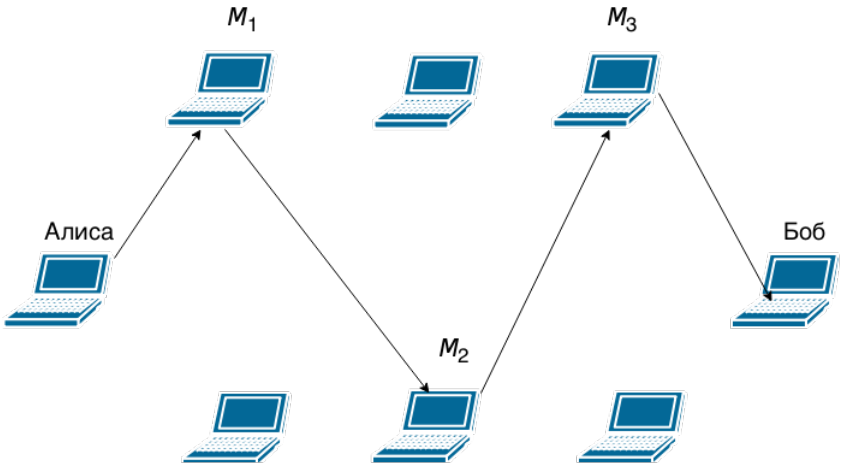


Рис. 2. Пример луковой маршрутизации сети Tor

Анонимность источника запроса обеспечивается за счёт того, что с точки зрения посредников сообщения (1) и (2) не отличаются. Если получатель, в свою очередь, тоже является участником сети Tor, то так же обеспечивается и его анонимность. Однако сеть Tor полагается на честность участников, что не подходит для конкурентной среды. Поэтому для решения поставленной задачи данный способ маршрутизации требует некоторых модификаций.

3.3. Безотказная луковая маршрутизация

Как сказано в п. 2, IP-адреса команд неизвестны, обращение к командам происходит по идентификаторам через посредников. Анонимность отправителя и получателя обеспечивается способом передачи сообщений и с помощью шифрования, которое защищает также целостность и конфиденциальность сообщения.

Для защиты доступности сообщений предлагается использовать принцип луковой маршрутизации сети Tor. Назовём этот способ *безотказной луковой маршрутизацией*.

Пусть Алиса отправляет Бобу сообщение m, M — посредник (рис. 3). Пример безотказной луковой маршрутизации проиллюстрирован на рис. 3. Для упрощения записи

будем считать, что E_X — это шифрование на открытом ключе пользователя X. Алиса шифрует сообщение следующим образом:

$$E_{Bob}(E_M(E_{Bob}(m))) \tag{3}$$

и отправляет получившуюся «луковицу» Бобу.

Получив сообщение (3), Боб расшифрует свой слой и передаст посреднику M следующее сообщение:

$$E_M(E_{Bob}(m)). \tag{4}$$

Посредник, повторив действия Боба, отправит в сеть сообщение вида

$$E_{Bob}(m). \tag{5}$$

Когда сообщение (5) дойдёт до Боба, он его прочитает и сможет ответить аналогичным образом, выбрав в общем случае другого посредника.

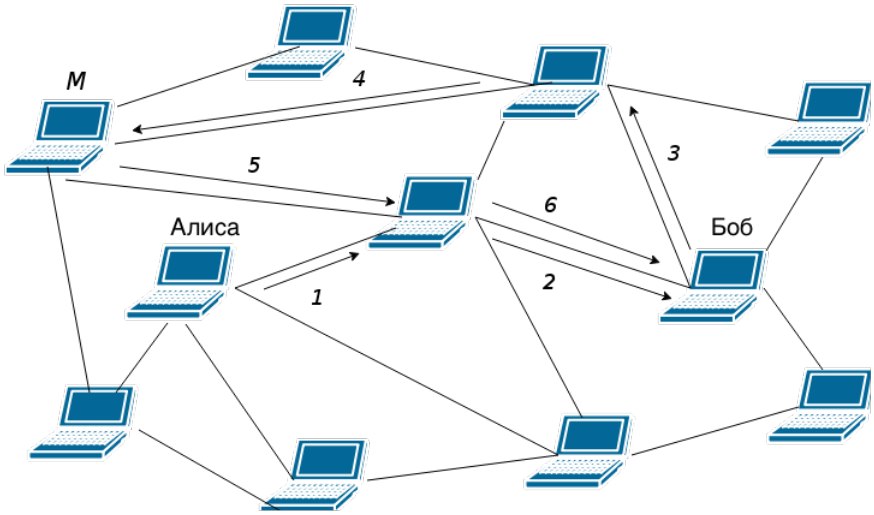


Рис. 3. Пример безотказной луковой маршрутизации

Рассмотрим, как обеспечивается невозможность отказа от посредничества. По полученному сообщению (3) Боб не сможет узнать ни источник сообщения, ни его получателя, а также не сможет определить, предназначено ли в конечном итоге это сообщение самому Бобу. Поэтому если Боб хочет получать сообщения из сети, ему придётся выступать в роли посредника для каждой полученной луковицы. С точки зрения любого посредника M, за исключением Боба, сообщения (3), (4) и (5) трактуются одинаково: «неизвестно, что внутри, но, возможно, это мне», поэтому сообщения передаются дальше. Посредники, не участвующие в образовании луковицы, вынуждены передавать сообщения по этой же причине.

Предложенный способ маршрутизации имеет существенное ограничение: он может быть применён только в тех случаях, когда каждый участник сети вынужден постоянно ожидать сообщения. Действительно, если по какой-либо причине участник сети уверен, что в данный момент не ожидает получения данных, то он может отбрасывать поступившие для передачи сообщения.

4. Протокол проведения соревнования

4.1. Обозначения и определения

Пусть sign — некоторый алгоритм цифровой подписи; verify — соответствующий ему алгоритм проверки подписи; $U\{ID_1, \ldots, ID_n\}$ — *множество идентификаторов* (открытых ключей) участников. Каждому открытому ключу ID_i соответствует закрытый ключ $\widehat{ID_i}$, $i \in \{1, \ldots, n\}$, то есть для любого сообщения x и любого $i \in \{1, \ldots, n\}$

$$\mathrm{verify}_{ID_i}(\mathrm{sign}_{\widehat{ID_i}}(x)) = 1.$$

Определим ответы на задания. Пусть $g_1, \ldots, g_m$ — открытые ключи (проверки подписи), $f_1, \ldots, f_m$ — соответствующие им закрытые ключи (подписания), то есть для любого $i \in \{1, \ldots, m\}$ и для любого сообщения x

$$\mathrm{verify}_{g_i}(\mathrm{sign}_{f_i}(x)) = 1.$$

Ключ f_i будем считать *ответом* на задание с номером i; g_i назовём *ключом проверки ответа* на задание с номером i.

Для того чтобы подтвердить получение ответа на задание в конкретное время конкретной командой, будем использовать метки времени.

Меткой времени от участника A для данных y, указывающей на момент времени t, будем считать набор значений $(t, y, \mathrm{sign}_{K_A}(t, y))$, где K_A — закрытый ключ (подписания) пользователя A.

Согласно требованиям к протоколу, нельзя полагаться на сервер выдачи меток времени, поэтому необходимо использовать распределённый подход — команды сами должны выдавать друг другу метки времени, подтверждающие получение ответа в конкретное время конкретным участником. За основу взят подход, описанный в [11]. Введём псевдослучайную функцию $G : \mathbb{N} \to U' \subset U$, $|U'| = t < n$.

Функция $G(y)$ используется в протоколе для определения множества участников, которые должны поставить метку времени под данными y. Для подтверждения получения данных y в указанное время требуется получить хотя бы k меток времени. Параметры n, t и k находятся в отношении $k < t < n$.

4.2. Протокол

Пусть на момент начала соревнования имеется сеть участников, описанная в п. 2, в которую команда организаторов входит как равноправный участник. Полагаем, что в этой сети для передачи сообщений используется безотказная луковая маршрутизация, описанная в п. 3.3. Каждый участник на момент начала соревнования имеет следующее:

1) алгоритмы sign и verify;
2) алгоритм симметричного шифрования на ключе E_x;
3) множество идентификаторов участников U;
4) ключи проверки ответов $g_1, \ldots, g_m$;
5) псевдослучайную функцию $G(y)$ с параметрами k, t;
6) зашифрованный набор заданий.

Началом соревнования считается момент рассылки организаторами ключа для расшифрования списка заданий, после чего команда организаторов перестаёт принимать активное участие.

Пусть некоторый участник u с идентификатором ID_u получил ответ f на задание. Рассмотрим протокол, согласно которому должен действовать участник u:

1) $y = \mathrm{sign}_f(t_u, ID_u)$, t_u — текущее время пользователя u.
2) Для всех пользователей с идентификатором $id \in G(y)$:
   а) $u \to id : z = E_x(y, t_u, ID_u)$, x — сеансовый ключ;
   б) $id \to u : t_{id}, \mathrm{sign}_{\widehat{id}}(z, t_{id})$, t_{id} — текущее время пользователя с идентификатором id.
3) Каждому пользователю с идентификатором $id \in U$ выслать:
   t_u, ID_u, $y = \mathrm{sign}_f(t_u, ID_u)$, $z = E_x(y, t_u, ID_u)$, x, $\{t_i, \mathrm{sign}_{\widehat{id_i}}(z, t_i) : i \in G(y)\}$.

Соревнование завершается в установленное организаторами время. Таблица результатов, имеющаяся у организаторов в этот момент, считается итоговой.

4.3. Пояснения к протоколу

Функция $G(y)$ используется с целью сделать труднопредсказуемым множество U' допустимых для запроса меток времени участников. Избыток участников в множестве U' объясняется тем, что среди команд множества U' могут быть как неактивные участники, так и злоумышленники, которые не будут выдавать метки времени для подтверждения ответа. Таким образом создаётся «запас» в виде $(t - k)$ участников для каждого ответа.

На первом шаге значение y служит сразу для двух целей. Во-первых, y является аргументом функции $G(y)$ на шаге 2, то есть определяет множество пользователей U', от которых будут набираться метки времени для подтверждения ответа. Поскольку значение y зависит от ответа f, его невозможно посчитать заранее. Во-вторых, значение y зависит от идентификатора участника, решившего задание, а значит, это значение не может быть использовано другим участником.

Необходимость шифровать запрос на выдачу метки времени для подтверждения ответа (шаг 2а) возникает для того, чтобы исключить возможность отказывать конкретному участнику в выдаче метки времени. После публикации факта получения ответа на шаге 3 все участники смогут расшифровать запрос и провести необходимые проверки.

4.4. Алгоритм проверки ответа

При получении сообщения (в шаге 3) каждый участник должен выполнить ряд проверок, прежде чем добавить в свою таблицу результатов новую запись. Пусть t — текущее время, w — условленное временное окно, g — ключ проверки ответа f. Алгоритм проверки факта получения ответа f следующий:

1) расшифровать $D_x(z)$;
2) если $t \leqslant_w t_u$ или $\exists i\ (t \leqslant_w t_i)$, то проверка не пройдена;
3) если среди множества выдавших метку времени для подтверждения ответа существует идентификатор $id \notin G(y)$, то проверка не пройдена;
4) если $\mathrm{verify}_g(y) = 0$, то проверка не пройдена;
5) $j := 0$; для всех $i \in G(y)$ выполнить:
   если $\mathrm{verify}_{ID_i}(\mathrm{sign}_{\widehat{ID_i}}(z, t_i)) = 1$, то $j = j + 1$;
6) если $j < k$, то проверка не пройдена;
7) проверка пройдена.

5. Атаки и предлагаемые улучшения

5.1. Атака отказа от посредничества

Описание атаки

Атака реализует угрозу нарушения работоспособности системы, что с высокой вероятностью приводит к невозможности участия в соревновании всех участников. Для защиты от этой атаки используется безотказная луковая маршрутизация, однако не созданы условия для её применения (см. п. 3.3). Участник может отказываться передавать не предназначенные ему данные до тех пор, пока ему самому не потребуется ожидать подтверждения факта получения ответа.

Противодействие

Предлагается наполнить сеть трафиком с помощью рассылки сообщений следующего вида:

$$(t, id_B, \operatorname{sign}_{\widehat{id_A}}(t, id_B)). \tag{6}$$

Сообщение (6) назовём *разрешением на запрос подтверждения получения ответа*, или просто *разрешением* от участника A участнику B, в котором t — это время создания разрешения, id_B — идентификатор участника B. Подразумевается, что разрешения рассылаются каждым участником каждому участнику.

К шагу 2а протокола следует добавить отправку разрешения $(t, ID_u, \operatorname{sign}_{\widehat{id}}(t, ID_u))$ от пользователя с идентификатором id пользователю u. Пользователь с идентификатором id выдаёт подтверждение только при наличии разрешения. Разрешение также должно быть в публикуемом факте получения ответа на шаге 3 протокола. Проверку наличия необходимых разрешений следует добавить в алгоритм проверки.

Таким образом, все участники постоянно будут ожидать получения разрешения от всех пользователей, так как предсказать, какие именно разрешения пригодятся, невозможно. Для того чтобы необходимость получения разрешений не прекращалась, можно ввести время жизни разрешения, то есть считать устаревшие разрешения недействительными.

5.2. Сговор с целью невыдачи метки времени для подтверждения ответа

Описание атаки

Атака реализует угрозу нарушения работоспособности системы, что с высокой вероятностью приводит к невозможности участия в соревновании конкретного участника либо всех участников. Участники не мотивированы выдавать метки времени для подтверждения чужих ответов (шаг 2б протокола). Если для некоторого $G(y)$ найдутся хотя бы $(t - k - 1)$ участников, которые отказались выдавать метку времени, то факт получения ответа не будет сформирован. Такая ситуация может сложиться не только в результате сговора злоумышленников, но и в результате отключения команд по иным причинам.

Противодействие

Протокол полагается на то, что большинство участников действует добросовестно. Пусть f — это количество злоумышленников в сети, $f < n$. Вероятность того, что

данная атака будет успешной, следующая:

$$p = 1 - \frac{\sum\limits_{i=0}^{t-k+1} \binom{n-f}{k+i}\binom{f}{t-k-i}}{\binom{n}{t}}.$$

В табл. 1 приведены вероятности для $t = [n/2]$, $k = [t/2]$, $n \in \{50, 100, 250, 500\}$.

Таблица 1

Вероятности успешной атаки п. 5.2

$n=50, t=25, k=12$		$n=100, t=50, k=25$		$n=250, t=125, k=62$		$n=500, t=250, k=125$	
f	p	f	p	f	p	f	p
14	0,0000	26	0,0000	69	0,0000	136	0,0000
15	0,0001	28	0,0000	74	0,0000	146	0,0000
16	0,0003	30	0,0000	79	0,0000	156	0,0000
17	0,0011	32	0,0000	84	0,0000	166	0,0000
18	0,0036	34	0,0001	89	0,0000	176	0,0000
19	0,0093	36	0,0008	94	0,0000	186	0,0000
20	0,0210	38	0,0035	99	0,0001	196	0,0000
21	0,0423	40	0,0121	104	0,0015	206	0,0000
22	0,0768	42	0,0338	109	0,0107	216	0,0008
23	0,1281	44	0,0791	114	0,0493	226	0,0123
24	0,1981	46	0,1579	119	0,1555	236	0,0895
25	0,2861	48	0,2742	124	0,3522	246	0,3274

5.3. Сговор с целью выдачи ложной метки времени конкретному участнику

Описание атаки

Атака может быть применена с целью нарушения работоспособности системы, что с высокой вероятностью приводит к искажению как собственных результатов, так и результатов другого конкретного участника. Если в множестве $G(y)$ окажется меньше k активных честных участников и хотя бы k участников вступят в сговор, то они смогут выдать ложную метку времени.

Противодействие

Как и в предыдущем случае, противодействие данной атаке полагается на то, что большинство участников сети — честные. Вероятность того, что атака окажется успешной по отношению к некоторому участнику, равна

$$p = \frac{\sum\limits_{i=0}^{t-k+1} \binom{f}{k+i}\binom{n-f}{t-k-i}}{\binom{n}{t}}.$$

В табл. 2 приведены вероятности для $t = [n/2]$, $k = [t/2]$, $n \in \{50, 100, 250, 500\}$.

5.4. Прочие атаки

Поскольку команды представлены лишь своим идентификатором, злоумышленник, помимо своей команды, может зарегистрировать несколько фиктивных и устроить сговор самостоятельно. Стоит заметить, что фиктивные команды могут вводиться и честными участниками с целью «уравновесить» злоумышленников. Можно пойти по пути

Таблица 2

Вероятности успешной атаки п. 5.3

$n = 50, t = 25, k = 12$		$n = 100, t = 50, k = 25$		$n = 250, t = 125, k = 62$		$n = 500, t = 250, k = 125$	
f	p	f	p	f	p	f	p
14	0,0018	26	0,0000	69	0,0000	136	0,0000
15	0,0061	28	0,0000	74	0,0000	146	0,0000
16	0,0161	30	0,0000	79	0,0000	156	0,0000
17	0,0359	32	0,0001	84	0,0000	166	0,0000
18	0,0699	34	0,0007	89	0,0000	176	0,0000
19	0,1218	36	0,0032	94	0,0001	186	0,0000
20	0,1934	38	0,0114	99	0,0009	196	0,0000
21	0,2836	40	0,0328	104	0,0073	206	0,0000
22	0,3881	42	0,0779	109	0,0370	216	0,0014
23	0,5000	44	0,1569	114	0,1265	226	0,0193
24	0,6112	46	0,2737	119	0,3063	236	0,1221
25	0,7139	48	0,4207	124	0,5503	246	0,3942

Bitcoin [12], в котором используется принцип защиты от злоупотребления услугами proof-of-work, предложенный в [13], и сделать участие в соревновании зависимым от решения сложной вычислительной задачи.

Отказ от рассылки разрешений можно приравнять к отказу от выдачи метки времени для подтверждения. Сюда же можно отнести рассылку разрешений с неверным временем. Точно так же включение неправильного времени в подтверждение ответа (без сговора) можно отнести к невыдаче метки времени для подтверждения ответа.

Заключение

В работе предложен удовлетворяющий поставленным требованиям протокол для проведения соревнований CTF, основанных на решении заданий.

В продолжение решения задачи проведения децентрализованных соревнований планируется рассмотрение проблем подготовительного этапа, построения сети, а также доработка предложенного протокола таким образом, чтобы снизить зависимость успешности проведения соревнований от добросовестности участников. Впоследствии планируется обобщить полученные результаты и описать протокол для решения абстрактной задачи.

ЛИТЕРАТУРА

1. *Колегов Д. Н., Чернушенко Ю. Н.* О соревнованиях CTF по компьютерной безопасности // Прикладная дискретная математика. 2008. № 2. С. 81–83.
2. `https://ctftime.org/ctf-wtf/` — CTFtime.org / All about CTF (Capture The Flag). 2014.
3. `https://www.defcon.org/` — DEF CON Hacking Conference. 2014.
4. `http://www.phdays.ru/` — PHDays — Positive Hack Days. 2014.
5. `http://ructf.org/` — RuCTF. 2014.
6. *Анисеня Н. И., Стефанцов Д. А., Торгаева Т. А.* Сервис BlackBox для проведения соревнований по защите компьютерной информации Capture the Flag // Прикладная дискретная математика. Приложение. 2013. № 6. С. 52–56.
7. `http://blackbox.sibears.ru/` — Blackbox. 2014.
8. `http://sibctf.ru/` — SiBCTF. 2014.
9. *Clarke I., Sandberg O., Wiley B., and Hong T. W.* Freenet: a distributed anonymous information storage and retrieval system // Intern. Workshop on Designing Privacy Enhancing

Technologies: Design Issues in Anonymity and Unobservability. N.Y.: Springer Verlag, 2001. P. 46–66.

10. `https://www.torproject.org/` — Tor Project. 2014.
11. *Haber S. and Stornetta W. S.* How to time-stamp a digital document // J. Cryptology. 1991. No. 3. P. 99–111.
12. `http://bitcoin.org/` — Bitcoin — Open source P2P money. 2014.
13. *Dwork C. and Naor M.* Pricing via processing or combatting Junk Mail // Proc. CRYPTO'92. Berlin, Heidelberg: Springer, 1993. P. 139–147.

REFERENCES

1. *Kolegov D. N., Chernushenko Yu. N.* O sorevnovaniyakh CTF po komp'yuternoy bezopasnosti [About the CTF — computer security competitions]. Prikladnaya Diskretnaya Matematika. 2008, no. 2, pp. 81–83. (in Russian)
2. `https://ctftime.org/ctf-wtf/` — CTFtime.org / All about CTF (Capture The Flag). 2014.
3. `https://www.defcon.org/` — DEF CON Hacking Conference. 2014.
4. `http://www.phdays.ru/` — PHDays — Positive Hack Days. 2014.
5. `http://ructf.org/` — RuCTF. 2014.
6. *Anisenya N. I., Stefantsov D. A., Torgaeva T. A.* Servis BlackBox dlya provedeniya sorevnovaniy po zashchite komp'yuternoy informatsii Capture the Flag [The BlackBox service for hosting Capture The Flag computer security competitions]. Prikladnaya Diskretnaya Matematika. Prilozhenie, 2013, no. 6, pp. 52–56. (in Russian)
7. `http://blackbox.sibears.ru/` — Blackbox. 2014.
8. `http://sibctf.ru/` — SiBCTF. 2014.
9. *Clarke I., Sandberg O., Wiley B., and Hong T. W.* Freenet: a distributed anonymous information storage and retrieval system. Intern. Workshop on Designing Privacy Enhancing Technologies: Design Issues in Anonymity and Unobservability, N.Y., Springer Verlag, 2001, pp. 46–66.
10. `https://www.torproject.org/` — Tor Project. 2014.
11. *Haber S. and Stornetta W. S.* How to time-stamp a digital document. J. Cryptology, 1991, no. 3, pp. 99–111.
12. `http://bitcoin.org/` — Bitcoin — Open source P2P money. 2014.
13. *Dwork C. and Naor M.* Pricing via processing or combatting Junk Mail. Proc. CRYPTO'92. Berlin, Heidelberg, Springer, 1993, pp. 139–147.

УДК 004.94

СКРЫТЫЕ КАНАЛЫ ПО ВРЕМЕНИ НА ОСНОВЕ ЗАГОЛОВКОВ КЭШИРОВАНИЯ ПРОТОКОЛА HTTP

Д. Н. Колегов, О. В. Брославский, Н. Е. Олексов

Национальный исследовательский Томский государственный университет, г. Томск, Россия

Описано неизвестное ранее семейство скрытых каналов по времени протокола HTTP на основе заголовков кэширования, построены их общая схема функционирования и две граничные модели угроз, в рамках которых предложены реализации и экспериментальные оценки рассматриваемых скрытых каналов. Показано, каким образом скрытые каналы данного семейства могут быть реализованы в веб-браузерах. Проведено сравнение обнаруженных скрытых каналов по времени на основе заголовков кэширования протокола HTTP с другими известными механизмами реализации скрытых каналов по времени в веб-браузерах. Разработаны программные компоненты для инструментального средства анализа защищённости BeEF, определяющие возможность реализации рассматриваемых скрытых каналов по времени в информационной системе.

Ключевые слова: *компьютерная безопасность, анализ защищённости, HTTP, информационные потоки, скрытые каналы, безопасность веб-приложений, безопасность веб-браузеров, бот-сети.*

DOI 10.17223/20710410/28/8

COVERT TIMING CHANNELS OVER HTTP CACHE-CONTROL HEADERS

D. N. Kolegov, O. V. Broslavsky, N. E. Oleksov

National Research Tomsk State University, Tomsk, Russia

E-mail: d.n.kolegov@gmail.com, o.v.broslavsky@gmail.com, n.e.oleksov@gmail.com

We introduce and discuss a new family of timing covert channels based on HTTP cache headers. We propose a general scheme of the timing covert channels in terms of access control models and data flow diagrams and suggest two base threat models for them. We then consider peculiarities of program implementation of the timing covert channels and their bandwidth depending on a HTTP cache header, a threat model, a programming language (C, JavaScript, Python, Ruby), and an environment. Finally we provide the basic characteristics of the implemented covert channels in web browsers and BeEF.

Keywords: *computer security, HTTP, cache-control headers, covert channels, web application security, web browsers security, botnets.*

Введение

Впервые скрытые каналы были описаны как механизмы сокрытия вредоносных данных внутри разрешённых данных, передаваемых по некоторому коммуникационному каналу [1]. В соответствии с нормативными документами *скрытым каналом* (covert

channel) называется не предусмотренный разработчиком системы информационных технологий и автоматизированных систем коммуникационный канал, который может быть применён для нарушения политики безопасности [2, 3].

Если первоначально в компьютерной безопасности скрытые каналы рассматривались в основном как источник угрозы нарушения политики безопасности мандатного управления доступом типа LBAC [4], приводящей к реализации запрещённых информационных потоков от сущностей с высоким уровнем конфиденциальности к сущностям с низким уровнем конфиденциальности, то в настоящее время скрытые каналы рассматриваются в контексте более широкого класса угроз [2, 5, 6], включающего в себя угрозы:

— скрытой передачи вредоносных программ и данных на контролируемые нарушителем компоненты информационной системы;
— передачи нарушителем команд управления агентам для выполнения;
— организации коммуникационных каналов управления бот-сетями;
— скрытой передачи криптографических ключей и параметров функционирования.

Традиционно по механизму передачи информации скрытые каналы разделяют на скрытые каналы по памяти и скрытые каналы по времени. Как правило, *скрытые каналы по памяти* основаны на наличии памяти, в которую передающий субъект записывает данные и из которой принимающий субъект считывает эти же данные. *Скрытые каналы по времени*, как правило, предполагают, что передающий субъект на основе данных модулирует некоторый изменяющийся во времени процесс, а принимающий субъект в состоянии демодулировать передаваемые данные, наблюдая изменение этого процесса во времени. Считается, что обнаружение скрытых каналов по времени является более сложной задачей, чем обнаружение скрытых каналов по памяти.

Для иллюстрации скрытого канала по времени рассмотрим следующий классический пример. Пусть в операционной системе имеются два процесса s_1 и s_2, которые могут создавать «слушающий» сокет на некотором заранее выбранном порту. Если процессу s_1 необходимо передать бит, равный 1, процессу s_2, то процесс s_1 создает «слушающий» сокет, если же необходимо передать бит, равный 0, то процесс s_1 ничего не делает. Для получения данных от процесса s_1 процесс s_2 пытается создать слушающий сокет на том же порту. Если при создании сокета субъектом s_2 в системе возникла ошибка, то процессом s_1 был передан бит 1, иначе — бит 0. Таким образом, процессы s_1 и s_2 с помощью рассмотренного скрытого канала, использующего механизм сокетов, реализуют в системе информационный поток, который является разрешённым или запрещённым в зависимости от политики безопасности информационной системы.

В настоящее время веб-приложения являются неотъемлемой частью систем информационных технологий и автоматизированных систем управления. В связи с этим проблема анализа защищённости веб-приложений является одной из самых актуальных в практической компьютерной безопасности. Одной из особенностей веб-приложений является использование в качестве клиента веб-браузера. Современные веб-браузеры являются наиболее распространённым, универсальным и широкоиспользуемым, но в то же время уязвимым программным обеспечением [7], что позволяет их использовать в различного рода атаках на информационые системы. Основным механизмом безопасности веб-браузеров является механизм Same Origin Policy. Обход данного механизма позволяет злоумышленнику выполнить произвольный код Javascript в веб-браузере пользователя в контексте (origin) уязвимого веб-приложения. Таким образом, после захвата контроля над веб-браузером возникает задача передачи управляю-

щей (сигнальной) информации от серверов управления нарушителя к веб-браузерам пользователей. Одним из способов передачи такой информации могут быть и скрытые каналы. Данные коммуникационные каналы, как правило, используются в бот-сетях и во вредоносном программном обеспечении.

Основным протоколом веб-приложений является протокол HTTP. Перечислим основные механизмы и элементы, используемые для реализации скрытых каналов в протоколе HTTP [7, 8]:

— перенаправление (redirect);
— HTTP cookies;
— заголовок HTTP Referer;
— произвольные (custom) HTTP-заголовки;
— коды ответов;
— пути и параметры, передаваемые в URL.

Большинство известных скрытых каналов в протоколе HTTP являются скрытыми каналами по памяти и основаны на отправке HTTP-сообщений определённого типа с заданными параметрами, несущими полезную информацию в теле запроса или ответа (HTTP-туннели), или на применении стеганографических методов для сокрытия факта передачи информации в HTTP-сообщениях.

Например, скрытый канал по памяти от клиента к серверу может быть реализован через заголовок кэширования *If-Range* следующим образом: передаваемая клиентом информация представляется в шестнадцатеричном виде и отправляется в значении заголовка *If-Range*. Так, передаваемое сообщение «hello» будет передано в виде «If-Range: 120c7bL-32bL-68656c6c6fL», где «68656c6c6f» — передаваемое сообщение, а «120c7bL-32bL» — стандартные части значения *ETag*-заголовка.

Примером скрытого канала с использованием стеганографических методов в HTTP-заголовках является скрытый канал через заголовок *Accept-Language*, изначально предназначенный для сообщения веб-серверу о поддерживаемых клиентом естественных языках. Предположим, что значение заголовка *Accept-Language*, равное «*en*», будет интерпретироваться сервером как 0, а «*fr*» — как 1. Тогда клиент может передать на сервер значение 0x50 путём отправки заголовка *Accept-Language* со значением «$en, fr, en, fr, en, en, en, en$».

Вместе с тем использование стеганографических методов, как правило, меняет стандартные значения заголовков или структуру HTTP-сообщений и, как следствие, приводит к обнаружению скрытого канала средствами фильтрации и контроля. Кроме того, для таких скрытых каналов может потребоваться модификация веб-сервера, позволяющая корректно обрабатывать недопустимые с точки зрения протокола значения заголовков.

В данной работе предлагаются скрытые каналы, использующие HTTP-заголовки по их прямому назначению, не вносящие дополнительных изменений в структуру заголовков и в формат передаваемых данных, а потому неотличимые от потока разрешённых HTTP-сообщений веб-приложения.

Обнаружению скрытых каналов по времени в протоколе HTTP уделяется мало внимания. Более того, авторам не известны описанные в научной литературе скрытые каналы по времени для протокола HTTP. Скрытые каналы, предложенные в работах [7 – 9], по своей сути являются скрытыми каналами, использующими механизмы функционирования протоколов IP, TCP или DNS, доступные через API веб-браузера, но не механизмы и структуры данных самого протокола HTTP.

В сетевых протоколах принято взаимодействующих субъектов разделять на субъектов-клиентов (или просто клиентов), которые инициируют соединение и отправляют запросы, и субъектов-серверов (или просто серверов), которые ожидают соединения, производят обработку запроса и возвращают клиентам соответствующие ответы. Таким образом, скрытые каналы в сетевых протоколах могут быть дополнительно классифицированы по направлению передачи информации на каналы от сервера к клиенту и от клиента к серверу, а также по возможности передачи данных — на однонаправленные (unidirectional) и двунаправленные (bidirectional).

Поскольку в протоколе HTTP именно клиент является инициатором соединения, реализация скрытых каналов, передающих информацию от клиента к серверу, является более простой и эффективной. В данной работе, напротив, пойдёт речь о скрытых каналах по времени, реализуемых от сервера к клиенту в протоколе HTTP, ввиду их большей ценности для исследований.

1. Заголовки кэширования протокола HTTP

Заголовком в протоколе HTTP называется пара вида «имя : значение». В HTTP-сообщениях заголовки отделяются друг от друга пустой строкой и предназначены для передачи служебной информации между клиентом и сервером. Например, заголовок *User-Agent* сообщает веб-серверу информацию о типе и версии веб-клиента, а заголовок *Accept-Language* содержит информацию о поддерживаемых веб-клиентом языках. Заголовки протокола HTTP делятся на заголовки запросов и заголовки ответов.

Заголовки запроса — это заголовки, используемые веб-клиентом для сообщения веб-серверу дополнительной информации (например, заголовок *Accept-Encoding*) или для уточнения характера запрашиваемой информации (например, заголовок *Content-Range*). *Заголовки ответа* — это заголовки, используемые веб-сервером для описания ответа (например, заголовок *Content-Length*, содержащий длину тела ответа, или заголовок *Last-Modified*, содержащий дату последнего изменения документа).

Кэширование является одним из основных механизмов протокола HTTP [10]. Его цель — устранение необходимости повторной отправки запросов и ответов, содержащих уже имеющиеся данные. Кэширование позволяет сократить время повторной загрузки веб-страницы, что приводит к увеличению производительности и доступности веб-приложений. Заголовки кэширования протокола HTTP позволяют клиенту и серверу обменяться информацией о наличии и актуальности запрашиваемых ресурсов.

Рассмотрим основные заголовки кэширования протокола HTTP, используемые в данной работе. Заголовки ответа, сообщающие клиенту о состоянии запрошенного ресурса:

1) *Last-Modified* — содержит дату последнего изменения запрошенного клиентом ресурса в формате «Tue, 10 Apr 2014, 12:34:56 GMT».
2) *ETag* (*entity-tag*) — содержит идентификатор запрошенного клиентом ресурса. Среди наиболее распространенных веб-серверов только в Apache стандартизован алгоритм формирования заголовка *ETag* [11]. Значение *ETag* в соответствии с данным алгоритмом формируется из шестнадцатеричных значений идентификатора ресурса (*inode*), его размера и времени последнего изменения в формате *mtime*: «ETag: 120c7bL-32bL-4f86d4105ac62L».

Заголовки запроса, сообщающие веб-серверу условия, при выполнении которых ему необходимо отправить клиенту изменённые ресурсы:

1) *If-Modified-Since*: HTTP-date;
2) *If-Unmodified-Since*: HTTP-date;

3) *If-Match*: entity-tag;
4) *If-None-Match*: entity-tag;
5) *If-Range*: HTTP-date или entity-tag.

Заголовки кэширования *If-Modified-Since* и *If-None-Match*, отправленные в запросе, позволяют определить, изменился ли ресурс на сервере. Сервер в ответ на такой запрос сообщает при помощи кода состояния HTTP-сообщения об изменении (200 OK) или о неизменении соответствующего ресурса (304 Not Modified). При этом в первом случае веб-сервер отправляет новую версию ресурса, а во втором ответ веб-сервера содержит только HTTP-заголовки.

Вторая пара заголовков *If-Unmodified-Since* и *If-Match* действует наоборот, спрашивая сервер, не изменился ли ресурс, и получая в ответ сообщение об изменении (412 Precondition Failed) или неизменении (200 OK).

Заголовок *If-Range* обычно используется в условных GET-запросах вместе с заголовками *If-Unmodified-Since*, *If-Match* или *Range* и позволяет уменьшить количество запросов, необходимых для получения новой полной версии запрашиваемого ресурса.

Таким образом, путём изменения ресурсов, запрашиваемых веб-клиентом, и получения отправляемых веб-сервером ответов возможна скрытая передача данных. Для этого достаточно принять следующую трактовку: изменение некоторого ресурса с момента последнего обращения к нему означает передачу бита 1, а его неизменение — бита 0. Рассмотрим возможные варианты реализации таких скрытых каналов, основанных на заголовках кэширования протокола HTTP.

2. Механизм функционирования скрытых каналов по времени на основе заголовков кэширования HTTP

В теории компьютерной безопасности *информационным потоком* от сущности-источника к сущности-приемнику называется преобразование данных в сущности-приемнике, осуществляемое субъектами информационной системы в зависимости от данных в сущности-источнике [12]. Рассмотрим представление информационного потока, порождаемого скрытым каналом на основе заголовков кэширования протокола HTTP, в рамках субъектно-сущностных схем управления доступом [12] и диаграмм потоков данных [13] (рис. 1).

Введём следующие обозначения на основе [12]: e_1 — сущность-файл, расположенная на стороне нарушителя и доступная на чтение субъекту s_1; e_3 — сущность-ресурс, расположенная на стороне веб-сервера и доступная на запись субъекту s_1 через некоторые интерфейсы веб-сервера s_2; e_2 — HTTP-запрос; e_4 — HTTP-ответ; e_5 — сущность-файл, доступная на запись процессу s_3 на стороне веб-клиента.

Пусть субъект s_1 хочет передать данные субъекту s_3. Тогда в каждый момент времени s_1 считывает один бит данных из сущности e_1 и, в зависимости от значения считанного бита, осуществляет или не осуществляет доступ на запись к ресурсу e_3. Важной особенностью является тот факт, что субъект s_1 не обязательно должен располагаться на веб-сервере, важно, чтобы он имел возможность осуществлять изменение данных в сущности e_3 на основе данных из сущности e_1 посредством сущностей-запросов e_2. Субъект-процесс s_3 выполняет HTTP-запрос к ресурсу e_3 и после получения HTTP-ответа e_4, отображающего изменения сущности e_3, пишет в сущность e_5 соответствующий бит (например, 1, если веб-страница была изменена, и 0 в противном случае).

Рассмотренная схема реализации информационного потока позволяет выделить две граничные модели угроз.

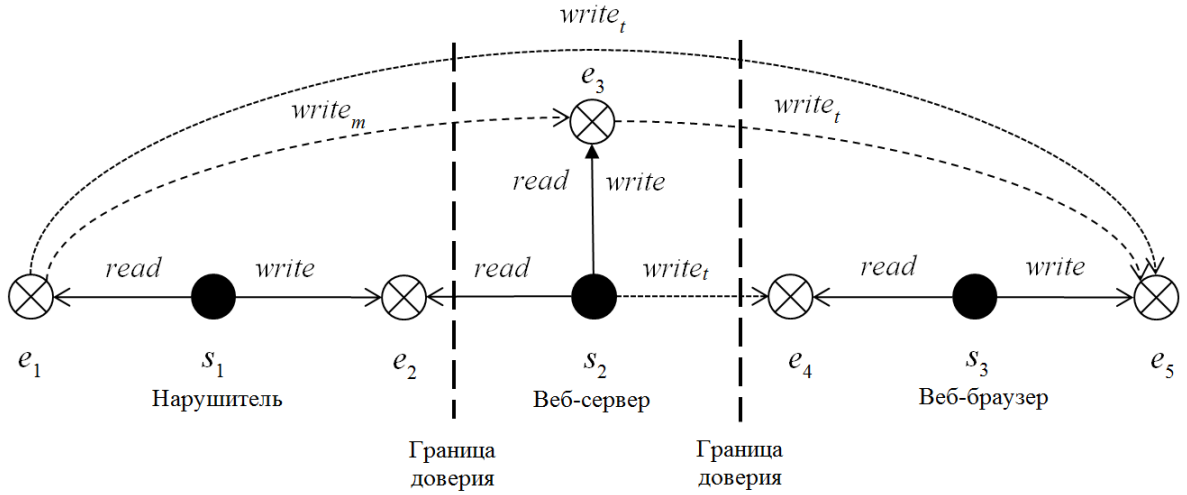


Рис. 1. Общая схема функционирования скрытых каналов по времени на основе заголовков кэширования HTTP

1) М о д е л ь M_1. Субъектами-нарушителями являются субъекты s_1 и s_3, субъект s_2 — доверенный веб-сервер (не контролируется нарушителем). Субъект-нарушитель s_1 является программным сценарием, имеющим возможность генерировать произвольные HTTP-запросы e_2 к интерфейсу веб-сервера s_2, зависящие от данных в e_1, а субъект-нарушитель s_3 является вредоносным сценарием JavaScript, функционирующим в веб-браузере на стороне пользователя. В рамках данной модели нарушитель не может напрямую изменять содержимое HTTP-ответа e_4, но, имея доступ на запись к ресурсу e_3, нарушитель может путём реализации информационных потоков по памяти от e_1 к e_3 менять время последней модификации ресурса e_3 и тем самым косвенно менять содержимое HTTP-ответа e_4. При этом субъект s_1 не имеет никакой информации о HTTP-запросах s_3 к s_2, то есть s_1 и s_3 функционируют независимо друг от друга, руководствуясь только временными интервалами. Данный интервал одинаков для процесса-отправителя и процесса-получателя и является параметром, выбираемым обеими сторонами на основании пропускной способности сети. В каждый такой временной интервал субъект s_1 осуществляет изменение сущности e_3, а субъект s_3, независимо от субъекта s_1, отправляет HTTP-запрос к e_3. В связи с этим возникает проблема синхронизации субъекта-отправителя s_1 и субъекта-получателя s_3. При этом неправильный выбор интервала приводит к увеличению ошибок при передаче в случае, когда время HTTP-запроса от s_3 к e_3 превышает выбранный интервал. Стоит отметить, что скрытый канал в модели M_1 обеспечивает анонимность для субъектов-нарушителей, т. е. средства контроля, функционирующие между границами доверия, не могут обнаружить сетевое взаимодействие субъектов-нарушителей s_1 и s_3.

2) М о д е л ь M_2. Субъектами-нарушителями являются субъекты s_1, s_2 и s_3. При этом субъект s_2 является веб-сервером, контролируется нарушителем и может произвольно модифицировать сущность-ответ e_4 на основе данных из e_3; s_3 — вредоносный сценарий JavaScript, функционирующий в веб-браузере на стороне пользователя. В данной модели взаимодействие субъектов-нарушителей s_1 и s_2 не имеет существенного значения. Для простоты можно считать, что в ней сущности e_1 и e_2 содержат одни и те же данные. Данный скрытый канал не обеспечивает свойства анонимности, но бо-

лее прост в реализации и обладает большей пропускной способностью. В этой модели также возникает задача обеспечения синхронизации, но уже между субъектами s_2 и s_3.

Рассматриваемые скрытые каналы по времени на основе заголовков кэширования протокола HTTP могут быть разделены на скрытые каналы на основе сущностей HTTP-date и на основе сущностей entity-tag. Отметим, что описанные выше модели угроз различаются лишь возможностями нарушителя по контролю веб-сервера и не накладывают никаких ограничений на способы модулирования полезной информации (payload) через заголовки кэширования протокола HTTP. Рассмотрим механизмы реализации, особенности и характеристики скрытых каналов каждой из групп в зависимости от используемой модели угроз.

3. Скрытые каналы по времени на основе сущностей HTTP-date

В механизмах кэширования протокола HTTP используются сущности HTTP-date в качестве меток времени. Например, заголовок *Last-Modified* содержит время последнего изменения ресурса, запрашиваемого клиентом: «Last-Modified: Tue, 10 Apr 2014, 12:34:56 GMT». Можно выделить следующие три варианта реализации скрытого канала по времени на основе сущностей HTTP-date.

1) На основе заголовка *Last-Modified*. Субъект s_3 обращается к сущности e_3 и получает HTTP-ответ e_4, содержащий начальное значение HTTP-date$_0$ заголовка *Last-Modified*. Для получения одного бита данных s_3 повторно обращается к сущности e_3 и сравнивает новое значение HTTP-date$_1$ заголовка *Last-Modified* со значением HTTP-date$_0$. Если значения в заголовках не совпадают, то субъектом s_1 была отправлена 1, а иначе 0.
2) На основе заголовка *If-Modified-Since*. Субъект s_3 обращается к сущности e_3 и получает HTTP-ответ e_4, содержащий начальное значение HTTP-date$_0$ заголовка *Last-Modified*. Затем субъект s_3 повторно обращается к сущности e_3 с заголовком «If-Modified-Since: HTTP-date$_0$» и получает HTTP-ответ e_4. Если код полученного ответа равен 200, то сущность e_3 была изменена и субъект s_1 отправил 1. Если код ответа равен 304, то сущность e_3 не изменялась и субъект s_1 отправил 0.
3) На основе заголовка *If-Unmodified-Since*. Субъект s_3 обращается к сущности e_3 и получает HTTP-ответ e_4, содержащий начальное значение HTTP-date$_0$ заголовка *Last-Modified*. Затем субъект s_3 повторно обращается к сущности e_3 с заголовком «If-Unmodified-Since: HTTP-date$_0$» и получает HTTP-ответ e_4. Если код полученного ответа равен 412, то сущность e_3 была изменена и субъект s_1 отправил 1. Если код ответа равен 200, то сущность e_3 не изменялась и субъект s_1 отправил 0.

Стоит отметить, что скрытые каналы на основе заголовков *If-Modified-Since* и *If-Unmodified-Since* возможны даже в случае, если на веб-сервере запрещено выставление заголовка *Last-Modified* в HTTP-ответе.

Заголовок *Last-Modified* хранит время последней модификации ресурса с точностью до секунды и обновляется веб-сервером, как правило, раз в секунду, поэтому максимальная пропускная способность скрытых каналов по времени на основе сущностей HTTP-date составляет 1 бит/с.

С целью оценки пропускной способности, достижимой на практике в сети интернет, реализован скрытый канал по времени на основе заголовка *Last-Modified* на языке программирования C с использованием библиотеки sys/socket.h. Скрытый канал реализован для модели угроз M_1: на веб-сервере субъект-нарушитель изменял запра-

шиваемый клиентом ресурс в зависимости от передаваемого бита данных. Реализация данного скрытого канала в рамках модели M_2 не может увеличить пропускную способность, однако позволяет повысить точность передачи ввиду того, что нет необходимости в синхронизации субъектов-нарушителей, поскольку в этом случае веб-сервер контролируется нарушителем и поэтому он может передавать следующий бит данных при поступлении запроса к ресурсу.

Введём следующие определения для оценки экспериментальных данных. *Минимальный корректный префикс* — это наименьшее количество бит, полученных от начала передачи данных и до первой ошибки. *Средний и максимальный корректные префиксы* являются количествами бит, переданных подряд без ошибок, в среднем и в лучшем случаях соответственно. *Точность передачи* — процентное отношение числа правильно полученных бит к общему числу переданных бит.

Результаты тестирования данной реализации представлены в табл. 1.

Таблица 1

Пропускная способность канала по времени на основе заголовка Last-Modified в модели M_1 в сети интернет

Интервал запроса, с	Мин. префикс, бит	Средний префикс, бит	Макс. префикс, бит	Скорость передачи, бит/с	Точность передачи, %
2	3400	10145	22143	0,5	99,87
1	3200	8848	19712	1	99,82

Таким образом, максимальная теоретическая пропускная способность скрытых каналов по времени на основе сущностей HTTP-date достижима в сети интернет.

4. Скрытые каналы по времени на основе сущностей entity-tag

Идентификатор entity-tag формируется, как правило, на основе данных *inode* документа в файловой системе, его размера и времени последнего изменения сущности [10, 11]. Можно выделить следующие три варианта реализации скрытого канала по времени на основе сущостей entity-tag.

1) На основе заголовка *ETag*. Субъект s_3 обращается к сущности e_3 и получает HTTP-ответ e_4, содержащий начальное значение entity-tag$_0$ заголовка $ETag$. Для получения одного бита данных s_3 повторно обращается к сущности e_3 и сравнивает новое значение entity-tag$_1$ заголовка $ETag$ со значением entity-tag$_0$. Если значения в заголовках не совпадают, то субъектом s_1 была отправлена 1, а иначе 0.
2) На основе заголовка *If-Match*. Субъект s_3 обращается к сущности s_3 и получает HTTP-ответ e_4, содержащий начальное значение entity-tag$_0$ заголовка $ETag$. Затем субъект s_3 повторно обращается к сущности e_3 с заголовком «If-Match: entity-tag$_0$» и получает HTTP-ответ e_4. Если код полученного ответа равен 412, то сущность e_3 была изменена и субъект s_1 отправил 1. Если код ответа 200, то сущность e_3 не изменялась и субъект s_1 отправил 0.
3) На основе заголовка *If-None-Match*. Субъект s_3 обращается к сущности e_3 и получает HTTP-ответ e_4, содержащий начальное значение entity-tag$_0$ заголовка $ETag$. Затем субъект s_3 повторно обращается к сущности e_3 с заголовком «If-None-Match: entity-tag$_0$» и получает HTTP-ответ e_4. Если код полученного ответа равен 200, то сущность e_3 была изменена и субъект s_1 отправил 1. Если код ответа 304, то сущность e_3 не изменялась и субъект s_1 отправил 0.

Скрытые каналы на основе заголовков *If-Match* и *If-None-Match* также возможно реализовать, даже если веб-сервер не выставляет заголовок *ETag*.

Поскольку сущности entity-tag, как правило, содержат время последней модификации ресурса в формате UNIX-time (количество микросекунд с 01.01.1970), то теоретическая пропускная способность данных скрытых каналов по времени равна 10^6 бит/с. Стоит отметить, что частота обновления заголовка не специфицирована в RFC и составляет, как правило, 1 с.

С целью оценки пропускной способности, достижимой на практике в сети интернет, реализован скрытый канал по времени на основе заголовка *ETag* на языке программирования C с использованием библиотеки sys/socket.h. Скрытый канал реализован для двух рассматриваемых моделей угроз. В рамках модели M_1 для сети интернет получены результаты, представленные в табл. 2.

Таблица 2

Пропускная способность канала по времени на основе заголовка ETag в модели M_1 в сети интернет

Интервал запроса, с	Мин. префикс, бит	Средний префикс, бит	Макс. префикс, бит	Скорость передачи, бит/с	Точность передачи, %
1	3200	8848	19712	1	99,82
0,5	2400	8142	18123	2	99,5

Для того чтобы максимально эффективно использовать пропускную способность данных скрытых каналов, необходимо, чтобы значение заголовка обновлялось при каждом изменении веб-сущности. В этом случае максимальная пропускная способность для каналов, основанных на заголовках *ETag*, составляет 1 бит за $(L+T)$ секунд, где L — время, необходимое s_3 для выполнения запроса к e_3 и получения ответа e_4, а T — время, необходимое субъектам s_2 и s_3 для вычислительных операций (сравнения значений заголовков, чтения и записи битов и т. п). Это может быть реализовано в модели нарушителя M_2. Предположения данной модели нарушителя обусловлены необходимостью изменить частоту выставления сервером заголовка *ETag*, чтобы максимально эффективно использовать предоставляемую им точность хранения времени. Для этого нарушитель должен иметь одну из следующих возможностей:

— изменять конфигурацию веб-сервера, увеличив частоту автоматического обновления заголовка *ETag*;
— изменять значение заголовка *ETag* в ответе HTTP после его формирования веб-сервером;
— использовать интерфейс выставления заголовков HTTP в динамических веб-страницах, предоставляемый данным веб-сервером, например, функцию header() языка PHP.

Таким образом, нарушитель должен иметь контроль над веб-сервером. Возможен также вариант использования модели M_1, если реализуется скрытый канал с использованием файлового хостинга, данный вариант рассмотрен далее. Важно, что даже при самостоятельном формировании значения заголовка *ETag* используется тот же алгоритм формирования значения заголовка (*inode-size-mtime*). Следовательно, сформированные нарушителем значения заголовков неотличимы от оригинальных ничем, кроме частоты обновления.

Для оценки пропускной способности в рамках модели угроз M_2 использовалась аналогичная реализация клиентской программы с использованием библиотеки sys/socket.h языка C. На стороне сервера использовалось веб-приложение, реализующее описанную выше логику работы на языке PHP. Полученные результаты представлены в табл. 3.

Таблица 3

Пропускная способность канала по времени на основе заголовка ETag в модели M_2 в компьютерных сетях различного типа

Тип компьютерной сети	Среднее значение HTTP RTT, мс	Скорость, бит/с
Узел сети	0,55	986
ЛВС ЦОД «*DigitalOcean*»	1,63	845,65
ЛВС	6,9	295,69
Интернет	110,2	13,22

Определим время отклика (HTTP round-trip time — RTT) как время, необходимое субъекту s_3 для того, чтобы сделать запрос к e_3 и получить ответ e_4 по протоколу HTTP. Пропускная способность рассматриваемых скрытых каналов существенно зависит от данного параметра, поэтому тестирование реализаций проводилось в компьютерных сетях различного типа:

— узел сети: субъекты s_1 и s_3 находятся на одном узле и общаются по протоколу HTTP;
— ЛВС ЦОД: субъекты s_1 и s_3 находятся на различных узлах одного ЦОД (в экспериментах использовался один из облачных ЦОДов «Digital Ocean»);
— ЛВС: s_1 и s_3 находятся на различных узлах одной ЛВС;
— интернет: s_1 и s_3 находятся на различных узлах, подключенных к сети интернет.

Кроме того, наложенные ограничения позволили s_2 изменять время последней модификации e_3 при формировании ответа на запрос s_3. Таким образом, появилась возможность отказаться от синхронизации интервалов между запросами s_1 и s_3 и избежать потерь данных в случаях, когда значение HTTP RTT для конкретного запроса превышало используемый интервал между запросами. Данное решение позволяет гарантировать 100 %-ю точность скрытого канала.

5. Реализация скрытых каналов по времени на основе заголовков кэширования в веб-браузерах

Реализация рассматриваемых скрытых каналов в веб-браузерах представляет отдельный интерес, поскольку последние являются основным клиентским приложением, использующим протокол HTTP, и в то же время накладывают существенные ограничения на используемые интерфейсы сетевого взаимодействия. Например, имеющимися средствами веб-браузеров в настоящее время невозможно получить доступ к низкоуровневым структурам IP- и TCP-заголовков сетевых пакетов, а также к сообщениям HTTP вне контекста (origin) веб-сервера. Вместе с тем, получив возможность исполнять произвольные сценарии Javascript в веб-браузере пользователя, нарушитель может запустить DDoS-атаку в отношении некоторого целевого сервера, получить доступ к конфиденциальной информации, доступной через интерфейс веб-браузера, выполнить сканирование внутренней локальной сети и т. д. Следовательно, возникает задача реализации коммуникационного канала для передачи команд и данных между серверами управления нарушителя и веб-браузером пользователя. Одним из способов

реализации таких коммуникационных каналов могут быть скрытые каналы. Таким образом, основная цель реализации скрытых каналов в веб-браузерах — это затруднение обнаружения и анализа сетевыми средствами контроля и фильтрации передаваемых данных между вредоносным сценарием Javascript, запущенным в веб-браузере пользователя (hooked browser), и сервером управления злоумышленника [14].

В настоящее время известные скрытые каналы по времени, допускающие реализацию в веб-браузерах, основаны на использовании допустимого времени доступа (таймеров) или DNS-туннелирования. Скрытый канал по времени на основе допустимого времени доступа предложен в [9] и реализован для протокола HTTP в [8]. Данный подход может быть использован для всех протоколов. Его идея заключается в следующем: два субъекта s_1 и s_2 договариваются о максимально возможном времени доступа t. Если субъект s_1 до истечения времени t получает запрос от s_2, то это означает передачу бита 1; если он не получает запроса, то был передан бит 0. Пропускная способность данного скрытого канала для протокола HTTP в ЛВС составляет 1,82 бит/с [8].

Скрытые каналы по времени на основе DNS-туннелирования [14, 15] реализуются следующим образом: веб-браузер отправляет DNS-запросы к серверу и интерпретирует ответы на них как последовательность бит. Например, сообщение об ошибке типа NXdomain на запрос $bit1.evil.com$ клиент интерпретирует как бит 0, а ответ на запрос $bit2.evil.com$, содержащий некоторый IP-адрес, как бит 1. Пропускная способность скрытого канала на основе DNS в сети интернет составляет примерно 10 бит/с.

Реализация предложенного в работе семейства скрытых каналов на основе заголовков кэширования в веб-браузере основана на использовании средств DOM, Javascript и HTML, накладывающих следующие дополнительные ограничения:

— невозможность синхронной остановки работы алгоритма на определённое время;
— низкая точность существующих асинхронных механизмов управления временем;
— сложность синхронизации субъекта-получателя и субъекта-отправителя;
— низкая точность передачи.

Недостижимость высокой синхронности работы кооперирующихся субъектов-нарушителей делает реализацию исследуемых скрытых каналов по времени нецелесообразной в рамках модели угроз M_1. Однако отказ от синхронизации, как в модели угроз M_2, позволяет обойти налагаемые средой веб-браузера ограничения и получить практическую реализацию скрытых каналов на основе заголовков кеширования протокола HTTP в веб-браузере.

Для оценки достижимой на практике пропускной способности исследуемых скрытых каналов по времени в веб-браузере разработаны клиентский сценарий, написанный на языке Javascript, и два серверных сценария, написанных на языках PHP и Python и размещённых на веб-серверах Apache и Flask соответственно. Полученные результаты представлены в табл. 4.

Таблица 4

Пропускная способность скрытых каналов по времени в веб-браузере

Заголовок	Версия сервера	HTTP RTT, мс	Скорость, бит/с
Last-Modified	Python	70	1
Last-Modified	PHP	68	1
ETag	Python	66	11,51
ETag	PHP	72	10,8

6. Реализация скрытого канала по времени с использованием облачных файловых хранилищ данных

Основным недостатком модели угроз M_1 является низкая пропускная способность скрытого канала, реализованного в её ограничениях. Как уже было сказано, наиболее распространённой частотой обновления заголовков кэширования является 1 с, а отсутствие контроля над веб-сервером s_2 (в том числе и невозможность модификации сущности e_4) в модели M_1 не позволяет нарушителю увеличить частоту обновления заголовка, а значит, и пропускную способность канала. В то же время использование в качестве s_2 веб-сервера с большей частотой обновления заголовков кэширования позволяет решить проблему низкой пропускной способности.

В качестве таких серверов возможно использование веб-серверов облачных файловых хранилищ данных. Основной целью облачных хостингов является размещение информации на сервере и обеспечение к ней доступа из сети интернет. Повышенная частота обновления заголовков кэширования обусловлена необходимостью поддержания актуальной информации о размещённых в облаке файлах, даже если они меняются чаще, чем раз в секунду. Кроме того, большинство крупных облачных хранилищ, например Dropbox или Google Drive, предоставляют программный интерфейс (API) для управления уже размещенными на сервере файлами: их загрузки и выгрузки, систематизации и обновления метаинформации (в том числе и времени последнего изменения файла). Таким образом, используя API, предоставляемые облачными хранилищами, можно реализовать скрытый канал на основе заголовков кэширования в рамках модели M_1, обеспечивающий свойство анонимности — субъекты s_1 и s_3 имеют возможность обмениваться данными, не взаимодействуя друг с другом напрямую.

В качестве платформы для тестирования выбран облачный сервис Google Drive. На хостинге был размещен файл e_3, доступный на запись субъекту s_1 и на чтение субъекту s_3; сами субъекты располагались на различных узлах сети интернет. Для передачи одного бита информации субъект s_1 совершает POST-запрос по адресу `https://www.googleapis.com/drive/v2/files/fileId/touch`, тем самым изменяя время последнего доступа к e_3 (здесь и далее fileId — идентификатор e_3 на облачном хостинге). Для получения переданного бита субъект s_3 получает значение используемого заголовка из метаинформации о файле, совершая GET-запрос по адресу `https://www.googleapis.com/drive/v2/files/fileId`, после чего восстанавливает переданный бит, руководствуясь описанными выше алгоритмами. Полученные данные представлены в табл. 5.

Таблица 5

Пропускная способность скрытого канала по времени с использованием Google Drive на основе *ETag*

Длина сообщения, бит	Точность, %	Скорость, бит/с	Ping RTT, мс
256	99,87	2,92	0,418
512	99,84	2,9	0,418
1024	99,8	2,88	0,418
2048	99,8	2,88	0,418
4096	99,87	2,86	0,418

Как видно из полученных результатов, пропускная способность скрытого канала в модели M_1 выросла и составляет теперь 1 бит за $(L + T + S)$ секунд, где L — время, необходимое s_3 для выполнения запроса к e_3; T — время, необходимое s_3 для вычислительных операций; S — время, уходящее на обработку запроса на серверах хостинга.

В данном случае время S выделяется отдельно, так как существенно зависит не от клиентов скрытого канала, а от выбранного для общения хостинга; при тестировании данное время составляло около 300 мс.

Таким образом, описанный скрытый канал сочетает в себе лучшие качества модели M_1, а именно: низкие требования; анонимность клиентов и трудность обнаружения и анализа сетевыми средствами контроля и фильтрации; кроме того, он обладает пропускной способностью, сравнимой с пропускной способностью скрытых каналов в модели угроз M_2. Высокий уровень доверия крупным облачным сервисам также играет на пользу данному скрытому каналу и затрудняет его обнаружение в защищённых средах.

7. Реализация скрытого канала по времени на основе заголовка *ETag* в BeEF

Browser Exploatation Framework (BeEF) — это инструментальное средство анализа защищённости информационных систем, ориентированное на использование веб-браузеров и эксплуатацию их уязвимостей [14].

Реализация скрытого канала по времени на основе заголовка *ETag* в BeEF состоит из двух частей. Первая часть — это расширение BeEF, написанное на языке Ruby и выполняющее функции специализированного веб-сервера, реализующего логику выставления заголовка *ETag* в соответствии с описанным выше методом. Вторая часть — это модуль BeEF, написанный на языке Javascript и реализующий клиентскую часть скрытого канала.

Исходный код реализованного скрытого канала доступен в репозитории BeEF [16] в следующих директориях:

— `/beef/extensions/etag` — серверная часть;
— `/beef/modules/ipec/etag_client` — клиентская часть.

При тестировании пропускной способности скрытого канала получены результаты, представленные в табл. 6.

Таблица 6

Пропускная способность скрытого канала по времени на основе *ETag* в BeEF

Тип сети	Скорость передачи 256 бит, бит/с	Скорость передачи 1024 бит, бит/с	Ping RTT, мс	HTTP RTT, мс
Local host	10,11	9,9	0,045	0,6
ЛВС «Digital Ocean»	10,08	9,84	0,062	0,83
ЛВС	10,03	9,78	18,046	19,8
Интернет	5,09	4,97	176,09	360,6

Заключение

В работе впервые описано семейство скрытых каналов по времени на основе заголовков кэширования протокола HTTP, обладающих следующими свойствами:

— для их реализации не требуется изменения грамматики сообщений HTTP;
— доступ к ресурсу на чтение, через который реализуется скрытый канал, не является блокирующим доступом и не приводит к возникновению ошибок передачи данных;
— возможна реализация анонимного скрытого канала, не требующего модификации веб-сервера;
— простота реализации;

— возможность реализации в веб-браузере;
— сложность обнаружения.

В рамках двух рассмотренных моделей угроз предложены способы реализации скрытых каналов по времени на основе заголовков кэширования в веб-браузерах, Google Drive API и BeEF. Описанные скрытые каналы превосходят по пропускной способности известные ранее скрытые каналы по времени [7, 8, 15] для веб-браузеров. Данная работа вошла в престижный топ-10 хакерских атак 2014 г. по версии компании «WhiteHat Security» [17].

ЛИТЕРАТУРА

1. *Lampson B. W.* A note on the confinement problem // Comm. ACM. 1973. No. 16(10). P. 613–615.
2. ГОСТ Р 53113.1 – 2008 Информационная технология. Защита информационных технологий и автоматизированных систем от угроз информационной безопасности, реализуемых с использованием скрытых каналов. Ч. 1. Общие положения.
3. ГОСТ Р 53113.2 – 2009 Информационная технология. Защита информационных технологий и автоматизированных систем от угроз информационной безопасности, реализуемых с использованием скрытых каналов. Ч. 2. Рекомендации по организации защиты информации, информационных технологий.
4. Timing Channels. `http://www.multicians.org/timing-chn.html`
5. CWE-514. Covert Channel. `https://cwe.mitre.org/data/definitions/514.html`
6. CWE-385. Covert Timing Channel. `https://cwe.mitre.org/data/definitions/385.html`
7. *Alkorn W., Frichot C., and Orru M.* The Browser Hacker's Handbook. Indianapolis: John & Wiley Sons, 2014. 648 p.
8. *Brown E., Yuan B., Johnson D., and Lutz P.* Covert channels in the HTTP network protocol: Channel characterization and detecting Man-in-the-Middle attacks // Proc. 5th Intern. Conf. Inform. Warfare and Security. Ohio, USA, April 8–9. The Air Force Institute of Technology, 2010. P. 56–65.
9. *Cabuk S., Brodley C.E., and Shield C.* IP covert timing channels: design and detection // Proc. 11th ACM Conf. on Computer and Communication Security. Washington DC, USA, 2004. P. 178–187.
10. RFC 2616. Hypertext Transfer Protocol HTTP 1.1. `http://www.w3.org/Protocols/rfc2616/rfc2616-sec13.html`
11. Apache Core Features Documentation. FileETag Directive. `http://httpd.apache.org/docs/2.2/mod/core.html#fileetag`
12. *Девянин П. Н.* Модели безопасности компьютерных систем. Управление доступом и информационными потоками: учеб. пособие для вузов. 2-е изд., испр. и доп. М.: Горячая линия-Телеком, 2013. 338 с.
13. Application Threat Modelling. `https://www.owasp.org/index.php/Application_Threat_Modeling`
14. The Browser Exploitation Framework Project. `http://beefproject.com/`
15. *Born K.* Browser-Based Covert Data Exfiltration. `http://arxiv.org/ftp/arxiv/papers/1004/1004.4357.pdf`
16. Исходный код BeEF. `https://github.com/beefproject/beef`
17. Top 10 Web Hacking Techniques of 2014. `https://blog.whitehatsec.com/top-10-web-hacking-techniques-of-2014/`

REFERENCES

1. *Lampson B. W.* A note on the confinement problem. Comm. ACM, 1973, no. 16(10), pp. 613–615.
2. GOST R 53113.1 – 2008 Informatsionnaya tekhnologiya. Zashchita informatsionnykh tekhnologiy i avtomatizirovannykh sistem ot ugroz informatsionnoy bezopasnosti, realizuemykh s ispol'zovaniem skrytykh kanalov. Part 1. Obshchie polozheniya [Information Technology. Protection of Information Technology and Automated Systems from Threats to Information Security Implemented using Covert Channels. Part 1. General Provisions]. (in Russian)
3. GOST R 53113.2 – 2009 Informatsionnaya tekhnologiya. Zashchita informatsionnykh tekhnologiy i avtomatizirovannykh sistem ot ugroz informatsionnoy bezopasnosti, realizuemykh s ispol'zovaniem skrytykh kanalov. Part 2. Rekomendatsii po organizatsii zashchity informatsii, informatsionnykh tekhnologiy. [Information Technology. Protection of Information Technology and Automated Systems from Threats to Information Security Implemented using Covert Channels. Part 2. Recommendations on the Organization of Information Security and Information Technology]. (in Russian)
4. Timing Channels. `http://www.multicians.org/timing-chn.html`
5. CWE-514. Covert Channel. `https://cwe.mitre.org/data/definitions/514.html`
6. CWE-385. Covert Timing Channel. `https://cwe.mitre.org/data/definitions/385.html`
7. *Alkorn W., Frichot C., and Orru M.* The Browser Hacker's Handbook. Indianapolis, John & Wiley Sons, 2014. 648 p.
8. *Brown E., Yuan B., Johnson D., and Lutz P.* Covert channels in the HTTP network protocol: Channel characterization and detecting Man-in-the-Middle attacks. Proc. 5th Intern. Conf. Inform. Warfare and Security. Ohio, USA, April 8–9. The Air Force Institute of Technology, 2010, pp. 56–65.
9. *Cabuk S., Brodley C.E., and Shield C.* IP covert timing channels: design and detection. Proc. 11th ACM Conf. on Computer and Communication Security. Washington DC, USA, 2004, pp. 178–187.
10. RFC 2616. Hypertext Transfer Protocol HTTP 1.1. `http://www.w3.org/Protocols/rfc2616/rfc2616-sec13.html`
11. Apache Core Features Documentation. FileETag Directive. `http://httpd.apache.org/docs/2.2/mod/core.html#fileetag`
12. *Devyanin P. N.* Modeli bezopasnosti komp'yuternykh sistem. Upravlenie dostupom i informatsionnymi potokami: ucheb. posobie dlya vuzov, 2-e izd., ispr. i dop. [Models of the Computer Systems Security. Access and Information Flow Control]. Moscow, Goryachaya Liniya-Telekom Publ., 2013. 338 p. (in Russian)
13. Application Threat Modelling. `https://www.owasp.org/index.php/Application_Threat_Modeling`
14. The Browser Exploitation Framework Project. `http://beefproject.com/`
15. *Born K.* Browser-Based Covert Data Exfiltration. `http://arxiv.org/ftp/arxiv/papers/1004/1004.4357.pdf`
16. Source Code BeEF. `https://github.com/beefproject/beef`
17. Top 10 Web Hacking Techniques of 2014. `https://blog.whitehatsec.com/top-10-web-hacking-techniques-of-2014/`

ПРИКЛАДНАЯ ТЕОРИЯ ГРАФОВ

УДК 519.6

СВОЙСТВА МИНИМАЛЬНЫХ ПРИМИТИВНЫХ ОРГРАФОВ

В. М. Фомичев

Финансовый университет при Правительстве Российской Федерации, ООО «Код безопасности», г. Москва, Россия

При $n \geqslant 4$ доказано, что сложность определения всех n-вершинных минимальных примитивных орграфов, являющихся частью заданного примитивного n-вершинного орграфа Γ, совпадает со сложностью распознавания монотонной булевой функции от s переменных, где s — число дуг (i, j) в Γ, таких, что полустепень исхода вершины i и полустепень захода вершины j превышают 1. Установлено, что при $n \geqslant 4$ все примитивные n-вершинные орграфы с числом дуг $n + 1$ являются минимальными и имеются минимальные примитивные n-вершинные орграфы с числом дуг от $n+2$ до $2n-3$. Описаны минимальные примитивные n-вершинные орграфы с числом дуг $n + 1$ и $n + 2$.

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

DOI 10.17223/20710410/28/9

PROPERTIES OF MINIMAL PRIMITIVE DIGRAPHS

V. M. Fomichev

Financial University under the Government of the Russian Federation, Moscow, Russia

E-mail: fomichev@nm.ru

It is proved that, for $n \geqslant 4$, the complexity of the determination of all n-vertex minimal primitive digraphs, which are parts of a given n-vertex primitive digraph Γ, coincides with the complexity of the recognition of a monotone Boolean function in s variables where s is the number of arcs (i, j) in Γ such that the vertex i out-degree and the vertex j in-degree exceed 1. It is found that, for $n \geqslant 4$, all the primitive n-vertex digraphs with $n+1$ arcs are minimal graphs and there are minimal primitive n-vertex digraphs with the number of arcs from $n + 2$ to $2n - 3$. Minimal primitive n-vertex digraphs with $n + 1$ and $n + 2$ arcs are described.

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

Введение

Запишем основные обозначения, используемые в работе:
$\mathbb{N}$ — множество натуральных чисел, $n, m \in \mathbb{N}$;
$(a_1, \ldots, a_n)$ — наибольший общий делитель натуральных чисел $a_1, \ldots, a_n$;
2^S — булеан множества S;

$M_0(n)$ — множество всех квадратных 0,1-матриц порядка n;
$M_0^P(n)$ — множество всех примитивных матриц из $M_0(n)$;
$M_0^P(n,m)$ — множество всех матриц из $M_0^P(n)$ с числом единичных элементов m;
$\Gamma(n)$ — множество всех орграфов с n вершинами;
$\Gamma^P(n)$ — множество всех примитивных орграфов с n вершинами;
$\Gamma^P(n,m)$ — множество всех примитивных орграфов с n вершинами и m дугами;
M — матрица смежности вершин орграфа Γ;
$[i,j]$ — простой путь в орграфе Γ из вершины i в вершину j.

В коммуникативных системах для исследования связей между элементами применяется матрично-графовый подход. Система из n элементов описывается с помощью n-вершинного орграфа Γ (или матрицы смежности его вершин $M=(m_{ij})$), в котором дуга (i,j) имеется тогда и только тогда, когда в системе i-й элемент влияет определённым образом на j-й элемент. Например, от i-го элемента непосредственно передаются данные j-му элементу, или i-й элемент является переменной величиной, от которой зависит j-й элемент, и пр., $i,j\in\{1,\ldots,n\}$.

Сложность реализации системы характеризуется, в частности, числом связей (дуг орграфа Γ). В [1] введено понятие минимальной примитивной матрицы как матрицы, которая после замены любого положительного элемента нулём не является примитивной. В силу естественной биекции между 0,1-матрицами порядка n и n-вершинными орграфами примитивный орграф Γ минимальный, если любая n-вершинная часть графа Γ не является примитивным графом. Минимальные примитивные матрицы и орграфы представляют интерес с точки зрения экономной реализации коммуникативной системы.

В работе продолжено начатое в [1] исследование свойств минимальных примитивных матриц и орграфов. Рассматриваются орграфы без петель и параллельных дуг.

1. Сложность определения минимальных примитивных n-вершинных орграфов, являющихся частями примитивного n-вершинного орграфа Γ

Обозначим: $M_{\min}^P(n)$ — множество всех минимальных примитивных матриц порядка n; $\Gamma_{\min}^P(n)$ — множество всех минимальных примитивных n-вершинных орграфов, являющихся частями примитивного n-вершинного орграфа Γ.

Оценим по Шеннону (то есть для наилучшего алгоритма при наихудших входных данных) сложность определения $\Gamma_{\min}^P(n)$, где элементарная вычислительная операция есть проверка примитивности любого n-вершинного орграфа или любой 0,1-матрицы порядка n.

Отметим некоторые свойства минимальных примитивных матриц [1].

Утверждение 1. Матрицы $A,B\in M_0(n)$, сопряжённые в группе подстановочных матриц:

а) одновременно примитивные или непримитивные;

б) в случае примитивности одновременно минимальные или неминимальные.

Утверждение 2. Множество $M_0(n)$ образует решётку в смысле отношения частичного порядка $\leqslant$, где $A\leqslant B\Leftrightarrow a_{ij}\leqslant b_{ij}$ для любых $i,j\in\{1,..,n\}$, $A=(a_{ij})$, $B=(b_{ij})$. Множество примитивных матриц $M_0^P(n)$ есть верхняя подполурешётка решётки $M_0(n)$, а множество $M_{\min}^P(n)$ — антицепь, состоящая из всех минимальных элементов подполурешётки $M_0^P(n)$.

Далее используем язык теории графов. В n-вершинном орграфе Γ обозначим p_i и q_i соответственно полустепени захода и исхода вершины $i\in\{1,\ldots,n\}$.

Утверждение 3. Если (i,j) — дуга орграфа Γ, то $\min\{q_i, p_j\} \geqslant 1$.

Пусть E — множество дуг орграфа Γ, $W \subseteq E$. Обозначим через Γ^W часть орграфа Γ, полученную из Γ удалением множества дуг $E \setminus W$.

Утверждение 4. Если Γ и Γ^W — примитивные орграфы, то $\min\{q_i, p_j\} > 1$ для любой дуги $(i,j) \in E \setminus W$.

Доказательство. По утверждению 3 $\min\{q_i, p_j\} \geqslant 1$. Если $\min\{q_i, p_j\} = 1$ для некоторой дуги $(i,j) \in E \setminus W$, например $q_i = 1$, то из вершины i исходит единственная дуга (i,j). После её удаления вершина становится концевой. Следовательно, получается не сильносвязный и не примитивный орграф. Случай $p_j = 1$ доказывается аналогично. ■

Утверждение 5. Если $W \subseteq U$, то из примитивности орграфа Γ^W следует примитивность орграфа Γ^U и из непримитивности орграфа Γ^U следует непримитивность орграфа Γ^W.

Для примитивного орграфа Γ подмножество дуг U назовём тупиковым в E, если орграф Γ^U примитивный и для любого собственного подмножества W множества U орграф Γ^W не примитивный. Отсюда следует, что система всех тупиковых в E подмножеств образует антицепь в решётке 2^E и имеется биекция между множеством $\Gamma^P_{\min}(n)$ и антицепью тупиковых в E подмножеств.

В орграфе Γ обозначим через S подмножество всех дуг (i,j) со свойством $\min\{q_i, p_j\} > 1$. Согласно утверждению 4, множество дуг $E \setminus S$ принадлежит любому орграфу из $\Gamma^P_{\min}(n)$. Тем самым доказана следующая теорема.

Теорема 1. Всякое тупиковое в E подмножество содержит $E \setminus S$, и имеется биекция между множеством $\Gamma^P_{\min}(n)$ и антицепью подмножеств Z множества S, таких, что $(E \setminus S) \cup Z$ — тупиковое в E подмножество.

Следствие 1. Пусть $|S| = s$, тогда сложность (по Шеннону) определения $\Gamma^P_{\min}(n)$ равна $\binom{s}{\lfloor s/2 \rfloor} + \binom{s}{\lfloor s/2 \rfloor + 1}$; размер необходимой памяти — порядка 2^s битов.

Доказательство. Согласно утверждению 5, задача определения $\Gamma^P_{\min}(n)$ равносильна задаче распознавания монотонной булевой функции $f : 2^S \to \{0,1\}$, где $f(Z) = 1$ тогда и только тогда, когда $(E \setminus S) \cup Z$ — множество дуг примитивного орграфа. Сложность распознавания монотонной булевой функции от s переменных равна указанной величине [2, с. 83]. ■

2. Описание минимальных примитивных орграфов

При изучении минимальных примитивных матриц и графов возникают следующие вопросы.

- При каких m класс $\Gamma^P(n,m)$ состоит только из минимальных примитивных графов?
- При каких m класс $\Gamma^P(n,m)$ не содержит минимальных примитивных графов?
- Как для данных n и m описать все минимальные примитивные графы из $\Gamma^P(n,m)$?

Решению некоторых из этих вопросов посвящены следующие результаты.

Теорема 2. При $n \geqslant 3$:

а) $P(n, n+1) \subset P_{\min}(n)$;

б) $\Gamma^P(n,m)$ содержит неминимальные матрицы, $n+2 \leqslant m \leqslant n(n-1)$;

в) $\Gamma^P(n,m)$ содержит минимальные матрицы при $m = n+1, \ldots, \max\{n+1, 2n-3\}$.

Доказательство.

а) Примитивная матрица M порядка n не имеет нулевых строк и столбцов, значит, число единиц в матрице M не меньше n. Кроме того, любая подстановочная матрица не примитивна. Тогда число единиц в матрице M больше n и любая примитивная матрица из $P(n, n+1)$ минимальная. Заметим, что $P(n, n+1) \neq \varnothing$ при $n \geqslant 3$, пример — матрицы смежности вершин графов Виландта [3, с. 109].

б) Пусть Γ — граф Виландта с множеством вершин $\{0, 1, \ldots, n-1\}$ и с множеством дуг $\{(n-2, 0)\} \cup \{(i, (i+1) \bmod n) : i = 0, 1, \ldots, n-1\}$, $n \geqslant 3$. При любом пополнении множества дуг получаем примитивный неминимальный орграф с числом дуг m, где $n + 2 \leqslant m \leqslant n(n-1)$.

в) Для $n = 3$ и для графа Виландта с тремя вершинами теорема выполнена.

Пусть $n \geqslant 4$. Рассмотрим n-вершинный орграф Γ с m дугами, который состоит из объединения r контуров взаимно простых длин $l_1, \ldots, l_r$, где $r, l_1, \ldots, l_r > 1$. Пусть пересечение множеств вершин всех контуров состоит из единственной вершины (обозначим её i), а каждая из остальных вершин орграфа Γ принадлежит ровно одному из контуров. Тогда орграф Γ сильносвязный и примитивный в соответствии с универсальным критерием [4], так как $r > 1$ и $(l_1, \ldots, l_r) = 1$. При $l_1, \ldots, l_r > 1$ удаление любой дуги нарушает сильную связность орграфа Γ и, следовательно, примитивность. Значит, орграф Γ минимальный.

Определим, при каких n и m существует указанный орграф Γ. По условию $m = l_1 + \ldots + l_r$, полустепени захода и исхода всех вершин, кроме i, равны 1, а для вершины i полустепени равны r. По теореме Эйлера m равно полусумме полустепеней захода и исхода всех вершин орграфа Γ, то есть $m = r+n-1$. В орграфе Γ имеется как минимум два контура взаимно простых длин, следовательно, их наименьшие возможные длины равны 2 и 3 соответственно, отсюда $n \geqslant 4$ и $m \geqslant 5$. При $n \geqslant 4$ наибольшее число контуров r с заданными условиями равно $n-2$ (один контур длины 3 и $n-3$ контуров длины 2). Если r пробегает все значения от 2 до $n-2$, то m пробегает все значения от $n+1$ до $2n-3$. ■

В орграфе Γ конкатенацию путей $w = (u, \ldots, i)$ и $w' = (i, \ldots, v)$, то есть путь $(u, \ldots, i, \ldots, v)$, обозначим $w \cdot w'$; вершину i отождествим с простым путём (i, i) длины 0.

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

Доказательство. Необходимость. Пусть $\Gamma \in \Gamma^P(n, n+1)$, тогда Γ не имеет висячих вершин. Значит, в Γ имеется вершина i с полустепенью захода 2, а остальные вершины имеют полустепени захода 1, и имеется вершина j с полустепенью исхода 2, а остальные вершины имеют полустепени исхода 1, где не исключено $i = j$. Согласно утверждению 1, включение $\Gamma \in \Gamma^P(n, n+1)$ инвариантно относительно перенумерации вершин орграфа Γ, поэтому без ограничения общности положим $i = n$.

Если $j = i = n$, то $(p_n, q_n) = (2, 2)$ и $(p_s, q_s) = (1, 1)$ для $s = 1, \ldots, n-1$. Следовательно, Γ есть объединение двух простых контуров длины l и λ с единственной общей вершиной n. Тогда $q = 0$, $l + \lambda = n + 1$ и $(l, \lambda) = 1$ в соответствии с универсальным критерием примитивности орграфа.

Пусть $j \neq n$, тогда $(p_n, q_n) = (2, 1)$, $(p_j, q_j) = (1, 2)$ и $(p_s, q_s) = (1, 1)$ для $s = 1, \ldots, n-1$, $s \neq j$. Следовательно, Γ есть объединение простого пути $[i, j]$ длины $q > 0$ и двух простых путей $[j, i]_1$ и $[j, i]_2$ длин $l - q$ и $\lambda - q$ соответственно, где множества вер-

шин путей попарно не пересекаются, за исключением начальной и конечной вершин. Отсюда Γ есть объединение контуров $[i,j]\cdot[j,i]_1$ и $[i,j]\cdot[j,i]_2$ длин l и λ, общая часть которых есть путь $[i,j]$. Тогда $(l,\lambda)=1$ в соответствии с универсальным критерием примитивности орграфа. Число дуг в Γ есть сумма длин путей $[i,j]$, $[j,i]_1$ и $[j,i]_2$, то есть $l+\lambda-q=n+1$, где $q\leqslant n-2$. Необходимость доказана.

Достаточность. Пусть n-вершинный орграф Γ есть объединение двух простых контуров взаимно простых длин l и λ, общая часть которых есть простой путь длины q, где $l+\lambda-q=n+1$, $0\leqslant q\leqslant n-2$, в случае $q=0$ общая часть контуров есть вершина. Тогда в соответствии с универсальным критерием Γ примитивный, так как $(l,\lambda)=1$. Число дуг в Γ есть число различных дуг, составляющих контуры, то есть $m=l+\lambda-q=n+1$. Число вершин в Γ равно n. ■

Для n-вершинного орграфа Γ обозначим $n_{r,s}$ число вершин с полустепенью захода r и полустепенью исхода s, $0\leqslant r,s,n_{r,s}\leqslant n$. Таблицу положительных чисел $\{n_{r,s}\}$ при всех допустимых значениях r и s назовём степенной структурой орграфа Γ, обозначается $D(\Gamma)$. Таблицу $D(\Gamma)$ запишем в виде $D(\Gamma)=\{(r,s)^{n_{r,s}}\}$, при $n_{r,s}=0$ элемент таблицы опускается. Например, степенная структура контура K длины n имеет вид $D(K)=\{(1,1)^n\}$; степенная структура орграфов, рассмотренных в теореме 3, имеет вид $\{(1,1)^{n-1},(2,2)^1\}$ при $j=n$ и $\{(1,1)^{n-2},(1,2)^1,(2,1)^1\}$ при $j\neq n$.

Заметим, что если графы изоморфны, то их степенные структуры совпадают.

Пусть (i,j) — дуга графа $\Gamma\in\Gamma(n)$. Назовём 1-расширением графа Γ орграф $\Gamma^{(i,j)}$ из $\Gamma(n+1)$, полученный добавлением к графу Γ вершины $n+1$ и заменой дуги (i,j) на две дуги: $(i,n+1)$ и $(n+1,j)$. Если Γ_k есть k-расширение графа Γ, то 1-расширение графа Γ_k назовем $(k+1)$-расширением графа Γ, $k\in\mathbb{N}$.

Пусть $\Gamma\in\Gamma^P(n,m)$ и K^* — система контуров в Γ. Система контуров K^* называется примитивной (минимальной примитивной), если натянутый на K^* подграф является примитивным (минимальным примитивным). Дугу (i,j) графа Γ назовем K^*-изолированной, если (i,j) не принадлежит ни одному из контуров системы K^*.

Теорема 4. Если $\Gamma\in\Gamma^P(n,m)$ при некоторых натуральных n и m, K^* — примитивная (минимальная примитивная) система контуров в Γ и в Γ имеется K^*-изолированная дуга (i,j), то при любом натуральном k имеется орграф Γ_k из $\Gamma^P(n+k,m+k)$, являющийся k-расширением графа Γ и содержащий систему K^*. Если при этом орграф Γ минимальный, то имеется k-расширение Γ_k, являющееся минимальным примитивным графом.

Доказательство. Индукция по k. Пусть $k=1$. Заметим, что система K^* содержится не только в Γ, но и в 1-расширении $\Gamma^{(i,j)}$ графа Γ, так как построение $\Gamma^{(i,j)}$ с использованием K^*-изолированной дуги (i,j) не изменяет системы K^*. При этом дуги $(i,n+1)$ и $(n+1,j)$ графа $\Gamma^{(i,j)}$ являются K^*-изолированными. Тогда в соответствии с универсальным критерием примитивности орграф $\Gamma^{(i,j)}$ является примитивным.

Если орграф Γ минимальный, то система контуров K^* минимальная примитивная. Удаление из Γ дуги (i,j) нарушает сильную связность; в силу K^*-изолированности дуги (i,j) невозможно нарушить примитивность при сохранении сильной связности. В силу минимальности орграфа Γ удаление любой из остальных дуг нарушает в Γ либо сильную связность, либо примитивность. Аналогично удаление дуги $(i,n+1)$ (или дуги $(n+1,j)$) из орграфа $\Gamma^{(i,j)}$ нарушает его сильную связность, удаление любой из остальных дуг нарушает в $\Gamma^{(i,j)}$ либо сильную связность, либо примитивность. Значит, орграф $\Gamma^{(i,j)}$ из $\Gamma^P(n+k,m+k)$ является минимальным примитивным.

Пусть теорема доказана для k-расширения Γ_k орграфа Γ при натуральных числах $1, \ldots, k$. Так как Γ_k удовлетворяет тем же условиям, что и Γ, рассуждения при переходе от Γ_k к Γ_{k+1} можно повторить. ■

Теорема 5. Если минимальный примитивный орграф $\Gamma \in \Gamma^P(n, n+2)$, то $D(\Gamma)$ принадлежит одному из классов, перечисленных в таблице:

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

Доказательство. Если сильносвязный орграф $\Gamma \in \Gamma^P(n, n+2)$, то числа $n_{r,s}$ связаны системой двух диофантовых уравнений, где первое уравнение перечисляет удвоенное число дуг в Γ (в соответствии с теоремой Эйлера), а второе — число вершин в Γ:

$$2n_{1,1} + 3n_{1,2} + 3n_{2,1} + 4n_{1,3} + 4n_{2,2} + 4n_{3,1} + 5n_{1,4} + 5n_{2,3} + 5n_{3,2} + 5n_{4,1} + 6n_{1,5} + \ldots + nn_{n-1,1} = 2n + 4,$$

$$n_{1,1} + n_{1,2} + n_{2,1} + n_{1,3} + n_{2,2} + n_{3,1} + n_{1,4} + n_{2,3} + n_{3,2} + n_{4,1} + n_{1,5} + \ldots + n_{n-1,1} = n.$$

Вычитая из первого уравнения удвоенное второе, получаем

$$n_{1,2} + n_{2,1} + 2n_{1,3} + 2n_{2,2} + 2n_{3,1} + 3n_{1,4} + 3n_{2,3} + 3n_{3,2} + 3n_{4,1} + 4n_{1,5} + \ldots + 4n_{5,1} + 5n_{1,6} + \ldots + (n-2)n_{n-1,1} = 4. \quad (1)$$

Определим решения уравнения (1) относительно целых неотрицательных чисел $n_{r,s}$ и укажем примитивные графы без петель, соответствующие полученным решениям.

Заметим, что $n_{r,s} = 0$ при $r + s > 6$, иначе левая часть уравнения (1) больше правой части, следовательно, уравнение (1) равносильно следующему упрощённому уравнению:

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

Имеется 9 классов решений уравнения (2).

1-й к л а с с. Если $n_{3,3} = 1$, то из уравнения (2) имеем

$$n_{1,2} = n_{2,1} = n_{1,3} = n_{2,2} = n_{3,1} = n_{1,4} = n_{2,3} = n_{3,2} = n_{4,1} = n_{1,5} = n_{2,4} = n_{4,2} = n_{5,1} = 0.$$

В этом случае Γ есть объединение трёх контуров, пересечение множеств вершин которых состоит из единственной вершины, а любая другая вершина принадлежит только одному из контуров (рис. 1), то есть $D(\Gamma) = \{(1,1)^{n-1}, (3,3)^1\}$.

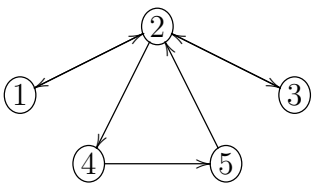


Рис. 1. Граф Γ, $n = 5$, $D(\Gamma) = \{(1,1)^4, (3,3)^1\}$

Для орграфа Γ, изображённого на рис. 1, $K^* = \{(1,2),(2,4,5)\}$, дуга $(2,3)$ является K^*-изолированной. Тогда по теореме 4 имеется k-расширение Γ_k орграфа Γ, где $D(\Gamma_k) = \{(1,1)^{k+4},(3,3)^1\}$, и из минимальности орграфа Γ следует минимальность орграфов Γ_k, $k \in \mathbb{N}$.

Если $n_{3,3} = 0$ и $n_{2,4} = 1$, то из уравнения (2) имеем

$$n_{1,2} = n_{2,1} = n_{1,3} = n_{2,2} = n_{3,1} = n_{1,4} = n_{2,3} = n_{3,2} = n_{4,1} = n_{1,5} = n_{4,2} = n_{5,1} = 0.$$

В этом случае в Γ имеется вершина i, где $p_i = 2$, $q_i = 4$, то есть имеются дуги (i,a), (i,b), (i,c), (i,d), где i,a,b,c,d различны. Орграф Γ сильносвязный, поэтому в Γ имеются простые пути $[a,i]$, $[b,i]$, $[c,i]$ и $[d,i]$. Так как $p_i = 2$, эти пути сходятся в два пути, то есть в Γ имеется вершина $j \neq i$, где $p_j \geqslant 2$. Тогда $n_{r,1} \geqslant 1$ при $r \geqslant 2$, то есть имеем противоречие.

Аналогичные противоречия получаем в следующих случаях:

а) $n_{3,3} = 0$ и $n_{1,5} = 1$, иначе в Γ имеется вершина, в которой сходятся от двух до четырёх путей, откуда следует $n_{2,1} + n_{3,1} + n_{4,1} \geqslant 1$;

б) $n_{3,3} = 0$ и $n_{5,1} = 1$, иначе в Γ имеется вершина, из которой расходятся от двух до четырёх путей, откуда следует $n_{1,2} + n_{1,3} + n_{1,4} \geqslant 1$;

в) $n_{3,3} = 0$ и $n_{4,2} = 1$, иначе в Γ имеется вершина, из которой расходятся не менее двух путей, то есть $n_{1,r} \geqslant 1$ при $r \geqslant 2$.

Значит, при $n_{3,3} = 0$ имеем $n_{1,5} = n_{2,4} = n_{4,2} = n_{5,1} = 0$ и уравнение (2) упрощается:

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

Из (3) следует, что не более чем одна из величин $n_{1,4}$, $n_{2,3}$, $n_{3,2}$, $n_{4,1}$ отлична от нуля.

2-й к л а с с. Если $n_{2,3} = 1$, то из уравнения (3) имеем $n_{1,3} = n_{2,2} = n_{3,1} = n_{1,4} = n_{3,2} = n_{4,1} = 0$, иначе левая часть уравнения (3) больше правой части. В этом случае в Γ имеется вершина i с полустепенью захода 2 и с полустепенью исхода 3, то есть имеются дуги (i,a), (i,b), (i,c), где i,a,b,c различны. Орграф Γ сильносвязный, поэтому в Γ имеются простые пути $[a,i]$, $[b,i]$ и $[c,i]$. Так как $p_i = 2$ и $n_{3,1} = 0$, то в Γ имеется вершина $j \neq i$, в которой сходятся два пути, то есть $n_{2,1} \geqslant 1$. Из уравнения (3) следует, что $n_{2,1} = 1$ и $n_{1,2} = 0$. Тогда Γ есть объединение трёх контуров, они пересекаются в единственной вершине i, и два контура сходятся в вершине $j \neq i$, то есть $D(\Gamma) = \{(1,1)^{n-2},(2,1)^1,(2,3)^1\}$.

Для орграфа Γ, изображённого на рис. 2, $K^* = \{(1,5),(1,3,4)\}$, дуга $(2,3)$ является K^*-изолированной. Тогда по теореме 4 имеется минимальное примитивное k-расширение Γ_k орграфа Γ, где $D(\Gamma_k) = \{(1,1)^{k+3},(2,1)^1,(2,3)^1\}$, $k \in \mathbb{N}$.

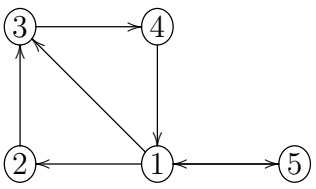


Рис. 2. Граф Γ, $n = 5$, $D(\Gamma) = \{(1,1)^3,(2,1)^1,(2,3)^1\}$

3-й к л а с с. Аналогично 2-му классу, при $n_{3,2} = 1$ имеем $n_{1,3} = n_{2,2} = n_{3,1} = n_{1,4} = n_{2,3} = n_{4,1} = 0$. Кроме того, $n_{1,2} = 1$ (из некоторой вершины расходятся два

пути) и $n_{2,1} = 0$. В этом случае Γ — объединение трёх контуров, пересечение множеств их вершин состоит из единственной вершины i, и два контура расходятся в вершине $j \neq i$, то есть $D(\Gamma) = \{(1,1)^{n-2}, (1,2)^1, (3,2)^1\}$.

Для орграфа Γ, изображённого на рис. 3, $K^* = \{(1,5), (1,4,3)\}$, дуга $(3,2)$ является K^*-изолированной. Тогда по теореме 4 имеется минимальное примитивное k-расширение Γ_k орграфа Γ, где $D(\Gamma_k) = \{(1,1)^{k+3}, (1,2)^1, (3,2)^1\}$, $k \in \mathbb{N}$.

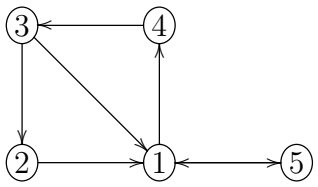


Рис. 3. Граф Γ, $n = 5$, $D(\Gamma) = \{(1,1)^3, (1,2)^1, (3,2)^1\}$

Если $n_{1,4} = 1$, то $n_{2,3} = n_{3,2} = n_{4,1} = 0$. Тогда в Γ имеется вершина i, где $p_i = 1$, $q_i = 4$, то есть имеются дуги (i,a), (i,b), (i,c), (i,d), где i, a, b, c, d различны. Орграф Γ сильносвязный, поэтому в Γ имеются простые пути $[a,i]$, $[b,i]$, $[c,i]$ и $[d,i]$. Так как $p_i = 1$ и $n_{4,1} = 0$, в Γ имеются либо две вершины, в которых сходятся соответственно два и три пути, либо три вершины, в которых сходятся по два пути. Следовательно, либо $n_{3,1} + n_{2,1} \geqslant 2$, либо $n_{2,1} \geqslant 3$. В обоих случаях левая часть уравнения (3) больше правой части.

Аналогичное противоречие получается при $n_{4,1} = 0$ и $n_{1,4} = n_{2,3} = n_{3,2} = 0$. Таким образом, при $n_{2,3} = n_{3,2} = 0$ имеем $n_{1,4} = n_{4,1} = 0$, и уравнение (3) равносильно упрощённому уравнению

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

Из (4) следует, что из величин $n_{1,3}$, $n_{2,2}$, $n_{3,1}$ не более чем две отличны от нуля и не более чем одна равна 2.

Если $n_{1,3} = 2$, то $n_{1,2} = n_{2,1} = n_{2,2} = n_{3,1} = 0$. Тогда в Γ имеется вершина i, где $p_i = 1$, $q_i = 3$, и дуги (i,a), (i,b), (i,c), где i, a, b, c различны, и простые пути $[a,i]$, $[b,i]$ и $[c,i]$. Так как $p_i = 1$ и $n_{3,1} = 0$, в Γ имеются две вершины, в которых сходятся по два пути. Следовательно, $n_{2,1} = 2$ — имеем противоречие. Аналогичные рассуждения отвергают следующие случаи:

— $n_{3,1} = 2$, $n_{1,2} = n_{2,1} = n_{1,3} = n_{2,2} = 0$;
— $n_{2,2} = n_{3,1} = 1$, $n_{1,2} = n_{2,1} = n_{1,3} = 0$;
— $n_{1,3} = n_{2,2} = 1$, $n_{1,2} = n_{2,1} = n_{3,1} = 0$;
— $n_{1,3} = 1$, $n_{1,2} = 2$, $n_{2,1} = n_{2,2} = n_{3,1} = 0$;
— $n_{3,1} = 1$, $n_{2,1} = 2$, $n_{1,2} = n_{1,3} = n_{2,2} = 0$;
— $n_{1,2} = 3$, $n_{2,1} = 1$, $n_{1,3} = n_{2,2} = n_{3,1} = 0$;
— $n_{1,2} = 1$, $n_{2,1} = 3$, $n_{1,3} = n_{2,2} = n_{3,1} = 0$.

В остальных случаях уравнение (4) имеет ещё шесть классов решений.

4-й к л а с с. Если $n_{2,2} = 2$, то $n_{1,2} = n_{2,1} = n_{1,3} = n_{3,1} = 0$ и $D(\Gamma) = \{(1,1)^{n-2}, (2,2)^2\}$.

Для орграфа Γ, изображённого на рис. 4, $K^* = \{(1,2,3), (1,2,3,4)\}$, дуга $(1,5)$ является K^*-изолированной. Тогда по теореме 4 имеется минимальное примитивное k-расширение Γ_k орграфа Γ, где $D(\Gamma_k) = \{(1,1)^{k+3}, (2,2)^2\}$, $k \in \mathbb{N}$.

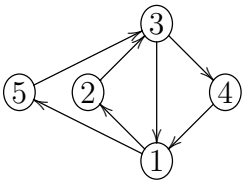


Рис. 4. Граф Γ, $n = 5$, $D(\Gamma) = \{(1,1)^3, (2,2)^2\}$

5-й к л а с с. Если $n_{1,3} = n_{3,}1 = 1$, то, согласно уравнению (4), $n_{1,2} = n_{2,1} = n_{2,2} = 0$ и $D(\Gamma) = \{(1,1)^{n-2}, (1,3)^1, (3,1)^1\}$.

Для орграфа Γ, изображённого на рис. 5, $K^* = \{(1,3), (1,2,3)\}$, дуга $(1,4)$ является K^*-изолированной. Тогда по теореме 4 имеется минимальное примитивное k-расширение Γ_k орграфа Γ, где $D(\Gamma_k) = \{(1,1)^{k+2}, (1,3)^1, (3,1)^1\}$, $k \in \mathbb{N}$.

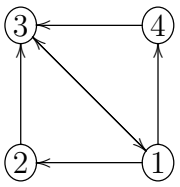


Рис. 5. Граф Γ, $n = 4$, $D(\Gamma) = \{(1,1)^2, (1,3)^1, (3,1)^1\}$

6-й к л а с с. Если $n_{1,3} = 1$, $n_{2,1} = 2$, то, согласно уравнению (4), $n_{1,2} = n_{3,1} = n_{2,2} = 0$ и $D(\Gamma) = \{(1,1)^{n-3}, (2,1)^2, (1,3)^1\}$.

Для орграфа Γ, изображённого на рис. 6, $K^* = \{(1,2,5), (1,2,3,4)\}$, дуга $(2,6)$ является K^*-изолированной. Тогда по теореме 4 имеется минимальное примитивное k-расширение Γ_k орграфа Γ, где $D(\Gamma_k) = \{(1,1)^{k+3}, (2,1)^2, (1,3)^1\}$, $k \in \mathbb{N}$.

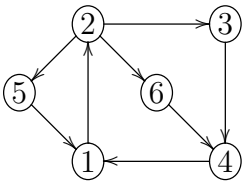


Рис. 6. Граф Γ, $n = 6$, $D(\Gamma) = \{(1,1)^3, (2,1)^2, (1,3)^1\}$

7-й к л а с с. Если $n_{3,1} = 1$, $n_{1,2} = 2$, то, согласно уравнению (4), $n_{2,1} = n_{2,2} = n_{1,3} = 0$ и $D(\Gamma) = \{(1,1)^{n-3}, (1,2)^2, (3,1)^1\}$.

Для орграфа Γ, изображённого на рис. 7, $K^* = \{(1,2,5), (1,2,3,4)\}$, дуга $(6,1)$ является K^*-изолированной. Тогда по теореме 4 имеется минимальное примитивное k-расширение Γ_k орграфа Γ, где $D(\Gamma_k) = \{(1,1)^{k+3}, (1,2)^2, (3,1)^1\}$, $k \in \mathbb{N}$.

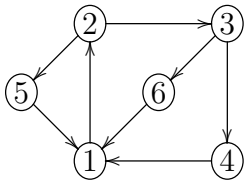


Рис. 7. Граф Γ, $n = 6$, $D(\Gamma) = \{(1,1)^3, (1,2)^2, (3,1)^1\}$

8-й к л а с с. Если $n_{2,2} = n_{1,2} = n_{2,1} = 1$, то, согласно уравнению (4), $n_{1,3} = n_{3,1} = 0$ и $D(\Gamma) = \{(1,1)^{n-3}, (1,2)^1, (2,1)^1, (2,2)^1\}$.

Для орграфа Γ, изображённого на рис. 8, $K^* = \{(1,2,3,4), (1,5,2,3,4)\}$, дуга $(6,4)$ является K^*-изолированной. Тогда по теореме 4 имеется минимальное примитивное k-расширение Γ_k орграфа Γ, где $D(\Gamma_k) = \{(1,1)^{k+3}, (1,2)^1, (2,1)^1, (2,2)^1\}$, $k \in \mathbb{N}$.

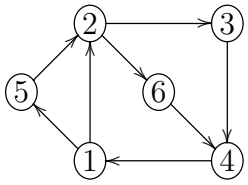


Рис. 8. Граф Γ, $n = 6$, $D(\Gamma) = \{(1,1)^3, (1,2)^1, (2,1)^1, (2,2)^1\}$

9-й к л а с с. Если $n_{1,2} = n_{2,1} = 2$, то, согласно уравнению (4), $n_{1,3} = n_{2,2} = n_{3,1} = 0$ и $D(\Gamma) = \{(1,1)^{n-4}, (1,2)^2, (2,1)^2\}$.

Для орграфа Γ, изображённого на рис. 9, $K^* = \{(1,2,5), (5,4,3,2)\}$, дуга $(3,6)$ является K^*-изолированной. Тогда по теореме 4 имеется минимальное примитивное k-расширение Γ_k орграфа Γ, где $D(\Gamma_k) = \{(1,1)^{k+2}, (1,2)^2, (2,1)^2\}$, $k \in \mathbb{N}$.

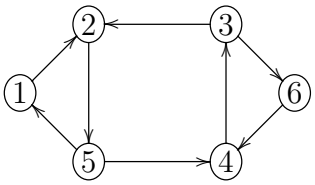


Рис. 9. Граф Γ, $n = 6$, $D(\Gamma) = \{(1,1)^2, (1,2)^2, (2,1)^2\}$

В силу полноты выполненного перебора вариантов не существует минимальных примитивных графов из $\Gamma^P(n, n+2)$ с другими степенными структурами. ■

ЛИТЕРАТУРА

1. *Бар-Гнар Р. И., Фомичев В. М.* О минимальных примитивных матрицах // Прикладная дискретная математика. Приложение. 2014. № 7. С. 7–9.
2. *Варфоломеев А. А., Фомичев В. М.* Информационная безопасность. Математические основы криптологии. Ч. I. М.: МИФИ, 1995. 114 с.
3. *Фомичев В. М.* Оценки экспонентов примитивных графов // Прикладная дискретная математика. 2011. № 2(12). С. 101–112.

4. *Харари Ф.* Теория графов. М.: Едиториал УРСС, 2003. 296 с.

REFERENCES

1. *Bar-Gnar R. I., Fomichev V. M.* O minimal'nykh primitivnykh matritsakh [About the minimal primitive matrices]. Prikladnaya Diskretnaya Matematika. Prilozhenie, 2014, no. 7, pp. 7–9. (in Russian)
2. *Varfolomeev A. A., Fomichev V. M.* Informatsionnaya Bezopasnost'. Matematicheskie Osnovy Kriptologii [Information Security. Mathematical Foundations of Cryptology]. Part I. Moscow, MEPhI Publ., 1995. 114 p. (in Russian)
3. *Fomichev V. M.* Otsenki eksponentov primitivnykh grafov [The estimates of exponents for primitive graphs]. Prikladnaya Diskretnaya Matematika, 2011, no. 2(12), pp. 101–112. (in Russian)
4. *Harary F.* Graph Theory. AW, 1969.

ВЫЧИСЛИТЕЛЬНЫЕ МЕТОДЫ В ДИСКРЕТНОЙ МАТЕМАТИКЕ

УДК 512.54.05+519.712.4

О СЛОЖНОСТИ ЗАДАЧИ ДИСКРЕТНОГО ЛОГАРИФМИРОВАНИЯ В ИНТЕРВАЛЕ В ГРУППЕ С ЭФФЕКТИВНЫМ ИНВЕРТИРОВАНИЕМ

М. В. Николаев

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

Задача дискретного логарифмирования в интервале заключается в поиске для заданной конечной группы $G = \langle P \rangle$ (с аддитивной записью операции) и заданных $P, Q \in G$, $N < |G| - 1$ такого значения n, что $Q = nP$, $n \in \{-N/2, \ldots, N/2\}$. Одним из наиболее эффективных методов решения данной задачи является алгоритм Годри — Шоста. В 2010 г. С. Гэлбрейт и Р. Руправи представили усовершенствованную версию алгоритма для групп с эффективным инвертированием. Оценка средней трудоёмкости решения задачи составила $(1{,}36 + o(1))\sqrt{N}$ групповых операций в G при $N \to \infty$. В настоящей работе приводится новая модификация алгоритма Годри — Шоста для решения задачи дискретного логарифмирования в интервале в группе с эффективным инвертированием и получена оценка средней трудоёмкости, составляющая $(1 + \varepsilon)\sqrt{\pi N/2}$ групповых операций в G.

Ключевые слова: *задача дискретного логарифмирования в интервале, алгоритм Годри — Шоста.*

DOI 10.17223/20710410/28/10

ON THE COMPLEXITY OF DISCRETE LOGARITHM PROBLEM IN AN INTERVAL IN A FINITE CYCLIC GROUP WITH EFFICIENT INVERSION

M. V. Nikolaev

Lomonosov Moscow State University, Moscow, Russia

E-mail: max.abstract@gmail.com

Discrete logarithm problem in an interval in a finite group $G = \langle P \rangle$ consists in solving the equation $Q = nP$ with respect to $n \in \{-N/2, \ldots, N/2\}$ for the specified $P, Q \in G$ and $0 < N < |G| - 1$. If the group G has an inversion, which may be computed significantly faster than the group operation, then, similarly to the solution of the classical discrete logarithm, we may speed up the algorithm. In 2010, S. Galbraith and R. Ruprai proposed an algorithm solving this problem with the average complexity $(1{,}36 + o(1))\sqrt{N}$ group operations in G where $N \to \infty$. We show that the average complexity of the algorithm for finding the solution of the discrete logarithm problem in interval equals $(1 + \varepsilon)\sqrt{\pi N/2}$ group operations.

Keywords: *discrete logarithm problem in interval, Gaudry — Schost algorithm.*

Приведём постановки задач.

Определение 1. Задача дискретного логарифмирования.
Дано: группа $G = \langle P \rangle$, $Q \in G$.
Найти: $n \in \{0, \ldots, |G| - 1\}$, такое, что $Q = nP$.

Определение 2. Задача дискретного логарифмирования в интервале.
Дано: группа $G = \langle P \rangle$, $Q \in G$, $N \in \mathbb{N}$, $2|N$, $N < |G| - 1$, $Q = nP$ для некоторого (неизвестного) $n \in \{-N/2, \ldots, N/2\}$.
Найти: n.

В настоящее время в общем случае одним из наиболее эффективным алгоритмом решения задачи дискретного логарифмирования в интервале является алгоритм Годри — Шоста [1]. Основная идея алгоритма может быть сформулирована следующим образом. Сначала выбираются так называемые «домашнее» (tame) и «дикое» (wild) множества:

$$T = \{-N/2, \ldots, N/2\}, \quad W = \{-N/2 + n, \ldots, N/2 + n\},$$

затем параллельно вычисляются псевдослучайные последовательности

$$x_i P, \ x_i \in T, \ i = 1, 2, \ldots, \tag{1}$$

$$Q + z_j P, \ (n + z_j) \in W, \ j = 1, 2, \ldots \tag{2}$$

до тех пор, пока в них не найдутся два одинаковых элемента

$$x_k P = Q + z_l P, \tag{3}$$

откуда находим $n = x_k - z_l$.

Средняя трудоёмкость алгоритма Годри — Шоста и его различных модификаций, измеряемая количеством групповых операций в G, равна по порядку величины среднему значению количества элементов последовательностей, вычисляемых до появления совпадающих элементов, в предположении, что значения n, x_i и z_j выбираются случайно равновероятно и независимо из соответствующих множеств. Это среднее значение может быть получено с использованием следующего результата Гэлбрэйта и Холмса, являющегося обобщением парадокса дней рождения.

Теорема 1 [2, Theorem 1]. Предположим, что выполнены следующие условия.

1) Имеется C различных цветов шаров, $C > 1$. Шар, выбранный под номером k, с вероятностью $r_{k,c}$ имеет цвет c (независимо от предыдущих выбранных шаров); для любого $c = 1, \ldots, C$ существует $p_c = \lim\limits_{n \to \infty} n^{-1} \sum\limits_{k=1}^{n} r_{k,c}$ и $p_1 \geqslant p_2 \geqslant \cdots \geqslant p_C > 0$. Пусть $b_{n,c} = p_c - n^{-1} \sum\limits_{k=1}^{n} r_{k,c}$ и существует константа K, такая, что для любого $c = 1, \ldots, C$, $n > 1$ выполняется неравенство $|b_{n,c}| \leqslant K/n$.
2) Имеется $N' \in \mathbb{N}$ различных урн. Если k-й шар имеет цвет c, то он попадает в урну с номером i с вероятностью $q_{c,i}(N')$ независимо от предыдущих выбранных цветов и размещений шаров. Существует такое $d > 0$, не зависящее от N' и c, что $0 \leqslant q_{c,i} \leqslant d/N'$ для любых $c = 1, \ldots, C$ и $i = 1, \ldots, N'$. Существуют такие константы $\alpha, \mu > 0$, что $|\{i \in \{1, \ldots, N'\} : q_{1,i}, q_{2,i} \geqslant \mu/N'\}| \geqslant \alpha N'$.

Тогда математическое ожидание числа $Z_{N'}$ шаров, размещенных до первого появления двух шаров разных цветов в одной урне, равно

$$\mathbf{M}(Z_{N'}) = \sqrt{\frac{\pi}{2A_{N'}}} + \mathrm{O}(N'^{1/4}),$$

где

$$A_{N'} = \sum_{c=1}^{C} p_c \left(\sum_{c'=1, c \neq c'}^{C} p_{c'} \left(\sum_{i=1}^{N'} q_{c,i} q_{c',i} \right) \right)$$

и константа в O зависит от $C, p_c, d, K, \alpha, \mu$, но не зависит от N' и $q_{c,i}$.

Для оптимизации методов поиска решения задачи дискретного логарифмирования часто используют наличие у исходной группы классов эквивалентности, как это делается, например, в работах [3, 4]. Но в случае задачи дискретного логарифмирования в интервале подойдут только классы эквивалентности, все элементы которых лежат в том же интервале, что и решение задачи. Итак, предположим теперь, что группа G обладает эффективно вычислимой операцией φ взятия обратного элемента, т. е. время, необходимое для вычисления обратного элемента, существенно меньше времени, необходимого для выполнения одной групповой операции. Тогда группа G распадается на непересекающиеся классы эквивалентности (орбиты) относительно действия φ, и подобно тому, как это делается в [4] для классической задачи дискретного логарифмирования, можно ускорить алгоритм, если искать не совпадающие элементы последовательностей (1) и (2), а совпадающие классы эквивалентности этих элементов. Действительно, в этом случае вместо равенства (3) имеем равенство

$$\varphi^s(x_k P) = Q + z_l P$$

для некоторого s, откуда $Q = ((-1)^s x_k - z_l)P$, т. е. $n = (-1)^s x_k - z_l$.

Примером такой группы *с эффективным инвертированием* является группа точек эллиптической кривой $y^2 = x^3 + Ax + B$ над конечным простым полем из $p > 3$ элементов. Действительно, $\varphi(x, y) = (x, -y)$, т. е. $\varphi(aP) = -aP$, и класс эквивалентности точки aP относительно действия группы $\langle\varphi\rangle$ состоит из aP и $\varphi(aP)$. Каждому такому классу эквивалентности соответствует множество $C(a) = \{a, -a\}$.

В [5] для этого случая предложена соответствующая модификация алгоритма Годри — Шоста, имеющая при $N \to \infty$ трудоёмкость $(1{,}36 + \mathrm{o}(1))\sqrt{N}$ групповых операций. Для получения этого результата использовались «домашнее» множество

$$T = \{C(a) : -N/2 \leqslant a \leqslant N/2\},$$

а также «дикое» множество

$$W = \{C(n + a) : -N/4 \leqslant a \leqslant N/4\}.$$

Используя описанный автоморфизм φ, получим, что множество «представителей» для каждого класса $C(a) \in T$ равно $\widetilde{T} = \{a : 0 \leqslant a \leqslant N/2\}$.

На рис. 1 изображены «дикое» множество, множество T_0 — объединение классов из множества T, а также пересечение $U = \widetilde{W} \cap \widetilde{T}$, где $\widetilde{W} = \{(n + a) : -N/4 \leqslant a \leqslant N/4\}$.

Следующая теорема конструктивно доказывает возможность дальнейшего улучшения оценки средней трудоёмкости решения задачи дискретного логарифмирования в интервале для группы с эффективным инвертированием.

Теорема 2. Пусть G — циклическая группа с эффективным инвертированием, пусть также $2|N$. Тогда для любого $\varepsilon > 0$ существует такой алгоритм решения задачи дискретного логарифмирования в интервале в группе G, что при случайном равновероятном выборе n его средняя трудоёмкость не превосходит $(1 + \varepsilon)\sqrt{\pi N/2} + \mathrm{O}_\varepsilon(N^{1/4})$ групповых операций, где $N \to \infty$. (Здесь запись O_ε означает, что константа под символом O зависит от ε.)

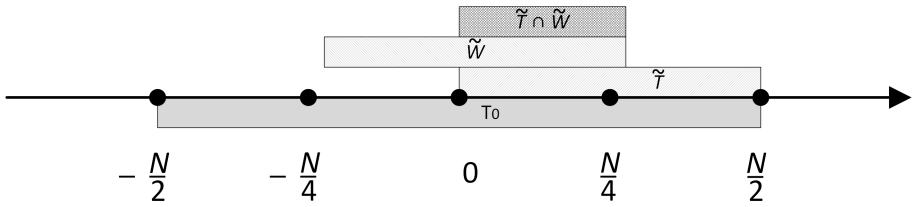


Рис. 1

Доказательство. Определим «домашнее» множество T и множество $\widetilde{T}$ представителей классов T

$$T = \{C(a) : -N/2 \leqslant a \leqslant N/2\}, \quad \widetilde{T} = \{0, \ldots, N/2\}$$

(в работах [5, 6] такое множество называется фундаментальной областью «домашнего» множества). Обозначим T_0 объединение классов из множества T. Для регулирования размера «дикого» множества W введём параметр τ:

$$W_\tau = \{C(n + a) : -\tau N/4 \leqslant a \leqslant \tau N/4\},$$

тогда $\widetilde{W}_\tau = \{a : -\tau N/4 + n \leqslant a \leqslant \tau N/4 + n\}$ и $|\widetilde{W}_\tau| = 2\lfloor \tau N/4 \rfloor + 1$.

Как и в алгоритме Годри — Шоста, будем параллельно вычислять последовательности точек

$$x_i P, \ x_i \in \widetilde{T}, \ i = 1, 2, \ldots, \tag{4}$$

$$Q + z_j P, \ z_j \in \widetilde{W}_\tau, \ j = 1, 2, \ldots \tag{5}$$

до тех пор, пока в них не найдутся две точки из одного класса эквивалентности, после чего находим решение задачи, как показано ранее. При этом предполагается, что значения x_i и z_j выбираются случайно равновероятно и независимо из соответствующих множеств.

Очевидно, что средняя трудоёмкость, выраженная в количестве групповых операций, не превосходит математического ожидания суммарного числа $Z_{N'}$ значений x_i и $(n+z_j)$, выбираемых до появления значений x_k и $(n+z_l)$, таких, что $C(x_k) = C(n_1+z_l)$.

Условное математическое ожидание $\mathbf{M}(Z_{N'}|(n))$ случайной величины $Z_{N'}$ при фиксированном n найдём с помощью теоремы 1, как это делается в [7]. Тогда в обозначениях теоремы $C = 2$; шары цвета 1 — элементы множества $\widetilde{T}$, а шары цвета 2 — элементы множества $\widetilde{W}_\tau$. Вычисление последовательностей (4) и (5) происходит параллельно, поэтому можно считать, что $r_{k,1} = r_{k,2} = 1/2$ для всех $k = 1, 2, \ldots$, откуда $p_1 = p_2 = 1/2$. Множество урн в нашем случае — это $T \cup W_\tau$, и шар a попадает в урну $C(a)$. Ясно, что $N' = \mathrm{O}(N)$. В целях упрощения записи далее величины $\mathrm{o}(N)$ при $N \to \infty$ опускаются. Тогда имеем

$$q_{1,i} = \begin{cases} 2/N, & \text{если } i \in T, \\ 0 & \text{в противном случае.} \end{cases}$$

С учётом последнего равенства и утверждения теоремы 1 нас интересуют значения $q_{2,i}$ только для $i \in T \cap W_\tau$. Поскольку каждый класс $C(a)$ содержит не более двух элементов, $T \cap W_\tau$ разбивается на два непересекающихся подмножества U_j, $j = 1, 2$, таких, что в каждый класс из U_j попадает ровно j элементов из $\widetilde{W}_\tau$, т. е. $q_{2,i} = j/|\widetilde{W}_\tau|$, $i \in U_j$.

Из двух последних равенств получаем

$$A_{N'} = 2 \cdot \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{2}{N} \cdot \frac{1}{|\widetilde{W}_\tau|} \sum_{j=1}^{2} j|U_j| = \frac{|U|}{N|\widetilde{W}_\tau|},$$

где $U = \widetilde{W}_\tau \cap T_0$, и по теореме 1 $\mathbf{M}(Z_{N'}|(n)) = \sqrt{\dfrac{\pi N|\widetilde{W}_\tau|}{2|U|}}$.

Следуя работам [5, 6], положим $n = xN/2$, $|x| \leqslant 1$. Оценим мощность множества U в зависимости от значения x.

1) $n \in B_1 = \{(xN) : |x| \leqslant 1 - \tau/2\}$ (рис. 2). Вероятность события $n \in B_1$ равна $(1 - \tau/2)$. В этом случае множество $\widetilde{W}_\tau$ полностью содержится в T_0, т. е. $|\widetilde{W}_\tau|/|U| = 1$.

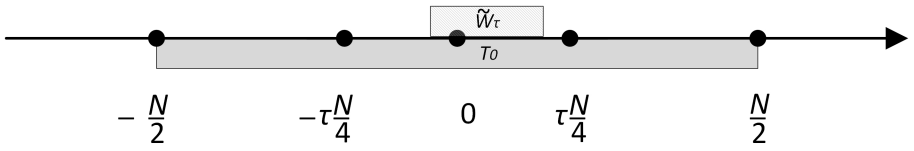


Рис. 2

2) $n \in B_2 = \{(xN) : |x| > 1 - \tau/2\}$ (рис. 3). Вероятность события $n \in B_2$ равна $\tau/2$. В этом случае можно сделать оценку $|\widetilde{W}_\tau|/|U| \leqslant 2$.

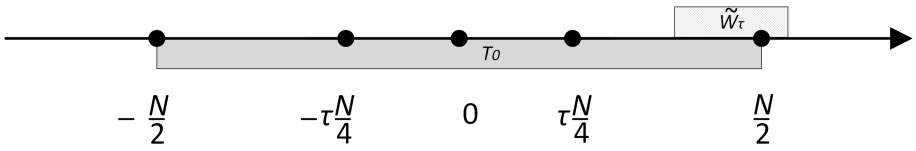


Рис. 3

Теперь можем оценить математическое ожидание:

$$\mathbf{M}(Z_{N'}) = \left(1 - \frac{\tau}{2}\right) \mathbf{M}(Z_{N'}|n \in B_1) + \frac{\tau}{2}\mathbf{M}(Z_{N'}|n \in B_2) \leqslant \left(1 - \frac{\tau}{2}\right)\sqrt{\pi N/2} + \frac{\tau}{2}\sqrt{\pi N} =$$
$$= \left(1 + \frac{(\sqrt{2} - 1)\tau}{2}\right)\sqrt{\pi N/2}.$$

Тогда при $\tau \to 0$ получаем утверждение теоремы. ■

ЛИТЕРАТУРА

1. *Gaudry P. and Schost E.* A low-memory parallel version of Matsuo, Chao and Tsujii's algorithm // LNCS. 2004. V. 3076. P. 208–222.
2. *Galbraith S. D. and Holmes M.* A non-uniform birthday problem with applications to discrete logarithms // Discr. Appl. Math. 2012. V. 160. No. 10–11. P. 1547–1560. `eprint.iacr.org/2010/616`
3. *Gallant R., Lambert R., and Vanstone S.* Faster point multiplication on elliptic curves with efficient endomorphisms // CRYPTO'2001. LNCS. 2001. V. 2139. P. 190–200.

4. *Wiener M. J. and Zuccherato R. J.* Faster attacks on elliptic curve cryptosystems // LNCS. 1999. V. 1556. P. 190–200.
5. *Galbraith S. D. and Ruprai R. S.* Using equivalence classes to accelerate solving the Discrete Logarithm Problem in a short interval // LNCS. 2010. V. 6056. P. 368–383. eprint.iacr.org/2010/615
6. *Liu W.* Improved algorithms for the 2-dimensional discrete logarithm problem with equivalence classes. MSc Thesis, University of Auckland, 2010. http://www.math.auckland.ac.nz/~sgal018/Wei-Liu-MSc.pdf
7. *Николаев М. В., Матюхин Д. В.* О сложности двумерной задачи дискретного логарифмирования в конечной циклической группе с эффективным автоморфизмом порядка 6 // Дискретная математика. 2013. Т. 25. № 4. С. 54–65.

REFERENCES

1. *Gaudry P. and Schost E.* A low-memory parallel version of Matsuo, Chao and Tsujii's algorithm. LNCS, 2004, vol. 3076, pp. 208–222.
2. *Galbraith S. D. and Holmes M.* A non-uniform birthday problem with applications to discrete logarithms. Discr. Appl. Math., 2012, vol. 160, no. 10–11, pp. 1547–1560. eprint.iacr.org/2010/616
3. *Gallant R., Lambert R., and Vanstone S.* Faster point multiplication on elliptic curves with efficient endomorphisms. CRYPTO'2001. LNCS, 2001, vol. 2139, pp. 190–200.
4. *Wiener M. J. and Zuccherato R. J.* Faster attacks on elliptic curve cryptosystems. LNCS, 1999, vol. 1556, pp. 190–200.
5. *Galbraith S. D. and Ruprai R. S.* Using equivalence classes to accelerate solving the Discrete Logarithm Problem in a short interval. LNCS, 2010, vol. 6056, pp. 368–383. eprint.iacr.org/2010/615
6. *Liu W.* Improved algorithms for the 2-dimensional discrete logarithm problem with equivalence classes. MSc Thesis, University of Auckland, 2010. http://www.math.auckland.ac.nz/~sgal018/Wei-Liu-MSc.pdf
7. *Nikolaev M. V. and Matyukhin D. V.* On the complexity of two-dimensional discrete logarithm problem in a finite cyclic group with effective automorphism of order 6. Discr. Math. Appl., 2013, vol. 23, iss. 3–4, pp. 313–326.

ДИСКРЕТНЫЕ МОДЕЛИ РЕАЛЬНЫХ ПРОЦЕССОВ

УДК 004.942

КЛЕТОЧНО-АВТОМАТНОЕ МОДЕЛИРОВАНИЕ ПРОЦЕССА РАЗРУШЕНИЯ ХРУПКИХ МАТЕРИАЛОВ

Д. В. Алексеев*, Г. А. Казунина**, А. В. Чередниченко**

*Кемеровский институт (филиал) Российского экономического университета им. Г. В. Плеханова, г. Кемерово, Россия

**Кузбасский государственный технический университет им. Т. Ф. Горбачева, г. Кемерово, Россия

Построен трёхмерный вероятностный клеточный автомат для моделирования эволюции кластерной структуры элементарных повреждений в нагруженных материалах. Проведено сравнение статистических характеристик временных рядов «число кластеров» и «число элементарных повреждений» трёхмерного и исследованного ранее двумерного клеточных автоматов. Показано, что переход временной автокорреляционной функции случайного процесса «число элементарных повреждений» в область отрицательных корреляций и появление второго линейного участка на статистике нормированного размаха Херста можно интерпретировать как предвестники перехода материала на стадию, непосредственно предшествующую необратимому разрушению. Установлено, что в трёхмерном случае существуют два качественно различных режима эволюции кластерной структуры, контролируемых вероятностью прорастания периметра кластера повреждений.

Ключевые слова: *клеточный автомат, кластеры элементарных повреждений, прогнозирование разрушения.*

DOI 10.17223/20710410/28/11

CELLULAR AUTOMATON SIMULATION OF THE FRACTURE PROCESS FOR BRITTLE MATERIALS

D. V. Alekseev*, G. A. Kazunina**, A. V. Cherednichenko**

*Kemerovo branch of Plekhanov Russian University of Economics, Kemerovo, Russia

**Kuzbass State Technical University named after T. F. Gorbachev, Kemerovo, Russia

E-mail: dmitriyalekseev@live.ru, gt-kga@yandex.ru, allacherednichenk@rambler.ru

A three-dimensional probabilistic cellular automaton is constructed to simulate the evolution of cluster structure of elementary damages in loaded materials. The comparison of the statistical characteristics of time series “number of clusters” and “number of elementary damages” are made for three-dimensional and two-dimensional cellular automata. It is shown, that the transition of the time autocorrelation function of a random process “number of elementary damages” to the range of negative correlations and the emergence of the second linear portion on the statistics of the normalized Hurst’s range can be interpreted as presages of material transition to the stage preceding to complete destruction. It is found that, for the three-dimensional model based

on the value of probability of damage cluster perimeter germination, there are two qualitatively different modes of damage accumulation.

Keywords: *cellular automaton, damage clusters, fracture prediction.*

Введение

Методы импульсной эмиссии (акустической и электромагнитной) используются в настоящее время как неразрушающие методы контроля прочности твёрдых материалов [1, 2]. При этом выявление тех параметров потока импульсной эмиссии, которые можно интерпретировать как предвестники разрушения материала, остаётся главной задачей и требует дополнительного исследования. Для прогнозирования разрушения главный интерес представляет пространственное распределение элементарных повреждений и их кластерная структура, тогда как характеристики импульсной эмиссии дают о ней только косвенную информацию [1, 2]. Одновременное наблюдение накопления повреждений и образуемой ими кластерной структуры в динамике на современном уровне развития технологии не представляется возможным. Поэтому является актуальным проведение подобного исследования методами компьютерного моделирования. В пользу такого подхода свидетельствует и наличие общих закономерностей на стадии, предшествующей разрушению материала [3]. По данным акустического эксперимента микротрещины, например, в горных породах образуются преимущественно на мезоскопическом уровне, и их средний размер находится в пределах $(1{,}4–28{,}4) \cdot 10^{-6}$ м. Поэтому процесс перехода разрушения на макроскопический уровень принципиально может быть описан без обращения к подробностям динамики отдельных элементарных актов, а с опорой только на геометрические характеристики рассматриваемой структуры. Простейшими моделями такого рода являются перколяционные модели, в рамках которых переход к макроскопическому разрушению описывается как геометрический фазовый переход.

Поскольку случайный процесс накопления повреждений в хрупких материалах является стохастическим, нелинейным и необратимым, адекватной математической моделью для описания этого процесса является модель вероятностного клеточного автомата, которая успешно используется для анализа сложных пространственно распределённых физико-химических процессов [4 – 6]. Так, в работах [7, 8] построена двумерная модель накопления повреждений и показано, что перед разрушением формируются степенные распределения дефектов по размерам, наличие которых является одним из признаков состояния самоорганизованной критичности. В работах [9, 10] разработана физическая концепция, математическая модель и комплекс программ для одновременного исследования кинетического процесса накопления повреждений и пространственно-временной эволюции их кластерной структуры в хрупких материалах при помощи нового двумерного вероятностного клеточного автомата. Проведённые модельные эксперименты позволили выявить параметры процесса накопления повреждений, характерные для неравновесных систем, склонных к катастрофам. Предложен новый качественный критерий перехода материала на стадию, непосредственно предшествующую разрушению, основанный на изломе нормированного размаха Херста и переходе выборочной временной корреляционной функции в отрицательную область. Настоящая работа является продолжением работ [9, 10] на трёхмерный случай и посвящена сопоставлению статистических характеристик временных рядов «число кластеров повреждений» и «число элементарных повреждений» для трёхмерного и двумерного клеточных автоматов.

1. Физическая концепция модели

Согласно кинетической теории прочности [1], процесс образования элементарных повреждений твёрдых материалов является термоактивационным стохастическим процессом. Поэтому для моделирования накопления повреждений предлагается вероятностный клеточный автомат, работа которого определяется набором вероятностей, характеризующих процессы образования элементарных повреждений по нескольким взаимодополняющим механизмам и генерирующих временные ряды числа элементарных повреждений и числа кластеров элементарных повреждений как результат эволюции пространственной кластерной структуры.

Вероятность образования нового элементарного повреждения на свободном узле решётки (вероятность оккупации) p_{occ} отображает интенсивность процесса разрушения материала под воздействием механических напряжений, усреднённых в пространственных масштабах, много больших характерного размера элементарного повреждения, и определяется внешними условиями нагружения материала. Вероятность прорастания периметра кластера повреждений p_{spr} отображает увеличенную интенсивность процесса разрушения материала под действием локальных перенапряжений вблизи уже имеющегося элементарного повреждения (кластера повреждений). Вероятность p_{mer} слияния кластеров, сблизившихся на критическое расстояние, учитывает взаимное влияние пары кластеров повреждений на их встречный рост. Разработанный клеточный автомат позволяет реализовывать различные сценарии моделирования накопления повреждений. В настоящей работе проводится сравнение однородного статического и динамического внутреннего сценариев, характеристики которых представлены в таблице.

Сценарии накопления повреждений

Сценарий	Вероятности, контролирующие процесс
Однородный статический	Постоянные вероятности оккупации p_{occ}, прорастания периметра p_{spr} и слияния кластеров p_{mer}
Динамический внутренний	Постоянные вероятности оккупации p_{occ} и слияния кластеров p_{mer}; зависящая от размера кластера вероятность прорастания периметра $p_{\text{spr}}(R) = p_{\text{spr}} \cdot e^{\frac{\gamma\sigma\sqrt[4]{R^2}}{kT\sqrt{l}}}$. Здесь $l^{-1/2}\sigma\sqrt[4]{R^2}$ — модельная оценка напряжения, где l — характерный размер элементарного повреждения [9, 10]

После выбора сценария моделирования и ввода входных параметров каждая итерация алгоритма генерации случайного процесса накопления повреждений работает в следующей последовательности. На каждой итерации работы автомата сначала образуются повреждения на неповреждённых узлах решётки, проращиваются периметры существующих кластеров; сливаются кластеры, сблизившиеся на критическое расстояние, проводится маркировка кластеров, что формирует кластерную структуру элементарных повреждений на данной итерации.

Конфигурация кластерной структуры на решётке на каждой итерации задаётся числом кластеров, а также характеристиками каждого кластера, такими, как масса (число элементов в кластере), среднеквадратичный радиус, размахи по строкам, столбцам, слоям. В результате каждая итерация даёт по одной точке в выборки временных рядов «число элементарных повреждений» и «число кластеров элементарных повреждений», по которым вычисляются характеристики этих временных рядов.

На каждой последующей итерации предыдущая кластерная структура заменяется новой кластерной структурой с автоматическим обновлением всех характеристик кластеров, то есть кластерная структура эволюционирует.

Конечной стадией эволюции кластерной структуры считается конфигурация, в которой образуется кластер, соединяющий противоположные грани куба. Образование соединяющего кластера интерпретируется как разрушение системы, а число итераций, необходимых для образования соединяющего кластера, отождествляется с временем до разрушения.

2. Формализация 3D-модели

Согласно алгоритму параллельных подстановок (АПП) [4], клеточный автомат (КА) представляем как композицию двух автоматов. Основной автомат №1 действует независимо, моделируя процесс разрушения системы, создаёт фрактальную кластерную структуру. Вспомогательный клеточный автомат №2 выполняет маркировку кластеров после каждой итерации работы основного автомата №1. Такая маркировка необходима для идентификации каждого отдельного кластера и описания кластерной структуры в целом.

Для формализации предложенной выше физической модели накопления повреждений необходимо различать внутренние (оккупированные) клетки и клетки внешнего периметра кластера повреждений. При этом внешний периметр (далее просто периметр) образуют клетки кластера повреждений, которые соприкасаются с неповреждёнными (неоккупированными) клетками. Поэтому алфавит основного КА, описывающий возможные состояния клетки, задаём как множество векторов из двух булевых компонент $u = (a, b)$:

$$A_1 = \{(0,0), (0,1), (1,0), (1,1)\}.$$

Здесь первый символ $a = 0$ ($a = 1$) означает, что клетка свободна (оккупирована). Второй символ $b = 0$ ($b = 1$) показывает, что клетка не принадлежит (принадлежит) периметру кластера повреждений. В итоге клетка в заданный момент времени может находиться в одном из следующих допустимых состояний u:

— (0,0) — клетка свободна и не принадлежит периметру;
— (1,0) — клетка оккупирована и не принадлежит периметру (внутренняя клетка кластера повреждений);
— (0,1) — клетка осталась свободной при проверке состояния клеток периметра кластера на предыдущей итерации («мёртвая клетка периметра»);
— (1,1) — клетка оккупирована и является клеткой периметра.

Задавая конечное множество имён (координат) клеток в дискретном пространстве как $M_1 = \{(i, j, k) : i, j, k = 0, 1, \ldots, n\}$, определяем шаблон соседства

$$T(i,j,k) = \{(i,j,k), \phi_1(i,j,k), \phi_2(i,j,k), \phi_3(i,j,k), \phi_4(i,j,k), \phi_5(i,j,k), \phi_6(i,j,k)\},$$

который указывает для каждой клетки ее ближайших соседей следующим образом:

$$\begin{aligned} \phi_1(i,j,k) &= (i-1,j,k); \qquad & \phi_2(i,j,k) &= (i+1,j,k);\\ \phi_3(i,j,k) &= (i,j-1,k); \qquad & \phi_4(i,j,k) &= (i,j+1,k);\\ \phi_5(i,j,k) &= (i,j,k-1); \qquad & \phi_6(i,j,k) &= (i,j,k+1). \end{aligned}$$

Таким образом, множество клеток

$$\Omega_1(t) = \{(u, (i,j,k)_1) : i, j, k = 0, 1, \ldots, n; u \in A_1\}$$

задаёт клеточный массив КА №1.

Согласно введённым определениям, кластером C_n называется множество клеток $((a,b),(i,j,k))$, каждая из которых имеет хотя бы одного соседа $\phi_d(i,j,k) \in T(i,j,k)$ в состоянии $u=(1,b)$. Если оккупированная клетка находится в состоянии $u=(1,1)$, а все соседи являются неповреждёнными и находятся в состоянии $u=(0,0)$, то клетка считается одиночным кластером и в то же время клеткой прорастающего периметра.

Состояния клетки в КА №1 изменяются под действием локальных операторов, соответствующих следующим действиям: $\theta_1(i,j,k)$ — оккупация клетки (образование нового элементарного повреждения), $\theta_2(i,j,k)$ — прорастание периметра кластера и $\theta_3(i,j,k)$ — слияние кластеров. Эти действия, согласно [4], записываются в общем виде как подстановка

$$\theta(i,j,k) : S(i,j,k) \cup S''(i,j,k) \xrightarrow{p} S'(i,j,k).$$

Здесь $S(i,j,k)$, $S''(i,j,k)$ и $S'(i,j,k)$ — локальные конфигурации:

— $S(i,j,k) = \{(v_0(i,j,k),(i,j,k)),\ldots,(v_d(i,j,k),\phi_d(i,j,k))\}$ — базовая конфигурация;
— $S''(i,j,k) = \{(w_0(i,j,k),\psi_0(i,j,k)),\ldots,(w_d(i,j,k),\psi_d(i,j,k))\}$ — контекст;
— $S'(i,j,k) = \{(v'_0(i,j,k),(i,j,k)),\ldots,(v'_d(i,j,k),\phi_d(i,j,k))\}$ — правая часть,

где $(v_0(i,j,k),(i,j,k))$ — узловая клетка подстановки S, принимающая допустимое состояние u; $v'_d = f_d(v_0,\ldots,v_d,w_0,\ldots,w_d)$ — функция переходов. Применение оператора $\theta(i,j,k)$ к клетке $(i,j,k) \in M_1$ состоит в замене значений клеток базовой конфигурации $S(i,j,k)$ на состояния клеток правой части $S'(i,j,k)$, вычисленных с помощью функции перехода v'_d.

Оператор $\theta_1(i,j,k)$ выполняет операцию оккупации свободных клеток с вероятностью оккупации p_{occ} с одновременным образованием кластеров из одной клетки при условии, что данная клетка окружена только свободными клетками:

$$\theta_1(i,j,k) : \{(0,0),(i,j,k)\} \cup \{((0,0),\phi_1(i,j,k)),\ldots,((0,0),\phi_6(i,j,k)),\} \xrightarrow{p_{\text{occ}}} \{v_0,(i,j,k)\},$$

где
$$v_0(i,j,k) = \begin{cases}(1,1), & \text{если } p_{\text{occ}} \leqslant \text{Rnd},\\ (0,0), & \text{если } p_{\text{occ}} > \text{Rnd},\end{cases}$$

Rnd — случайное число, равномерно распределённое в диапазоне [0,1].

Оператор $\theta_2(i,j,k)$ выполняет операцию проращивания кластеров по периметру, обновляя периметры кластеров следующим образом: ближайшие соседи клеток периметра уже существующих кластеров присоединяются к кластеру с вероятностью p_{spr}, которая в случае динамического внутреннего сценария моделирования зависит от размера кластера (среднеквадратичного радиуса). Другими словами, если свободная клетка $((0,0),(i,j,k))$ является соседней с клеткой периметра, то она с вероятностью p_{spr} переходит в состояние $u=(1,1)$, то есть становится клеткой нового периметра. Одновременно клетка периметра становится внутренней клеткой кластера, то есть получает состояние $u=(1,0)$. Но если при сканировании ближайших соседей клетка не перешла в состояние $u=(1,1)$, она становится «мёртвой клеткой периметра», переходя в состояние $u=(0,1)$. Заметим, что в дальнейшем росте кластера за счёт прорастания периметра участвуют только клетки в состоянии $u=(1,1)$. Поэтому оператор $\theta_2(i,j,k)$ имеет вид

$$\theta_2(i,j,k) : \{(0,0),(i,j,k)\} \cup \{(1,1),\phi_d(i,j,k)\} \xrightarrow{p_{\text{spr}}} \{(v_0,(i,j,k)),(v_d,\phi_d(i,j,k))\},$$

где $d \in \{1,\ldots,6\}$,

$$v_0(i,j,k) = \begin{cases}(1,1), & \text{если } p_{\text{spr}} \leqslant \text{Rnd},\\ (0,1), & \text{если } p_{\text{spr}} > \text{Rnd};\end{cases} \qquad v_d(i,j,k) = \begin{cases}(1,0), & \text{если } p_{\text{spr}} \leqslant \text{Rnd},\\ (1,1), & \text{если } p_{\text{spr}} > \text{Rnd}.\end{cases}$$

Алгоритм работы КА итерационный. Первой итерацией автомата №1, когда все клетки свободны, является оккупация свободных клеток клеточного массива $\Omega_1(t)$ с вероятностью p_{occ} путём применения оператора $\theta_1(i,j,k)$.

Непосредственно после этой итерации основного КА второй клеточный автомат №2 выполняет первую маркировку кластеров: присваивает каждой оккупированной клетке (состояния $u = (1,1)$) номер кластера, которому она принадлежит. Поэтому алфавит вспомогательного клеточного автомата задаётся как совокупность номеров $A_2 = \{1, 2, \ldots, N_{\max}\}$, где $N_{\max}$ — максимальное число кластеров, полученное в данной итерации работы автомата. Множество имён клеток вспомогательного автомата изоморфно множеству имён клеток основного автомата: $M_1 \sim M_2$. При этом второй КА №2 функционирует, используя КА №1 как контекст и создавая соответствующий порядок клеточного массива $(u, (i,j,k)_1) \in \Omega_1(t)$.

Маркировка кластеров выполняется согласно алгоритму, который подобен алгоритму роста кластеров Хамерсли — Лиса — Александровица [11]. Простейший вариант этого алгоритма на кубической решётке можно описать следующим образом. Создаётся «вспомогательный массив клеток», который в дальнейшем будет использован маркировкой; первоначальные значения его элементов соответствуют номерам оккупированных клеток.

Первая итерация работы КА №2 начинается сканированием клеточного массива $\Omega_1(t) = \{u, (i,j,k)_1 : i,j,k = 0,1,\ldots,n\}$ и при обнаружении оккупированной клетки (на первой итерации основного КА это состояние $u = (1,1)$) продолжается созданием кластера, то есть присвоением клетке выбранного кластера определённого номера:

$$\theta_4(i,j,k) : \{(1,1), (i,j,k)_1\} \cup \{((0,0), \phi_d(i,j,k)_1), ((0,1), \phi_l(i,j,k)_1)\} \to \{z, (i,j,k)_2\},$$

где $d, l \in \{1,\ldots,6\}$, причём $d \neq l$.

Заметим, что хотя на первой итерации выбранная оккупированная клетка окружена только свободными клетками (состояние $u = (0,0)$), в последующих итерациях такая первоначально выбранная клетка может иметь в соседях также и «мёртвые клетки периметра» (состояние $u = (0,1)$).

Следующий шаг — формирование начального периметра кластера путем присоединения к выбранной клетке соседних оккупированных клеток, то есть клеток, находящихся в состояниях $u = (1,1)$ или $u = (1,0)$, с присвоением всем клеткам выбранного номера кластера, что можно записать в виде подстановки

$$\theta_5(i,j,k) : \{z, (i,j,k)_2\} \cup \{((1,1), \phi_d(i,j,k)_1), ((1,0), \phi_l(i,j,k)_1)\} \to \{z, (i,j,k)_2\},$$

где $d, l \in \{1,\ldots,6\}$, причём $d \neq l$.

Такие итерации формирования периметра продолжаются до тех пор, пока новый периметр не оказывается пустым (состояния $u = (0,0)$, $u = (0,1)$). При этом все клетки кластера получают номер (первый).

Далее берётся другая оккупированная клетка решётки, не входящая в сформированный кластер, и повторяются все перечисленные итерации с присвоением клеткам кластера другого (второго) номера.

Маркировка заканчивается, как только все оккупированные клетки решётки оказываются присоединёнными к какому-либо кластеру.

На каждой итерации работы основного КА, начиная со второй, выполняются следующие действия:

1) Кластеры клеточного массива $\Omega_1(t)$ проращиваются по периметру при помощи оператора $\theta_2(i,j,k)$.

2) Оператор слияния $\theta_3(i,j,k)$ соединяет пару кластеров, разделённых одним узлом (клеткой). Согласно построенной модели, вероятность оккупации свободной клетки p_{mer} в этом случае принимает значение, увеличенное по сравнению с вероятностью p_{occ}. Действие оператора слияния $\theta_3(i,j,k)$ происходит следующим образом: если клетка периметра (состояние $u=(1,1)$) имеет в своём соседстве клетку периметра, принадлежащую другому кластеру, то с вероятностью p_{mer} обе клетки становятся клетками одного кластера. Обозначая номера разных кластеров X и Y, $X \neq Y$, получим

$$\theta_3(i,j,k) : \{(0,0),(i,j,k)_1\} \cup \{((1,1),\phi_d(i,j,k)_1,$$
$$(X,\phi_d(i,j,k)_2)),((1,1),\phi_l(i,j,k)_1,(Y,\phi_l(i,j,k)_2))\} \xrightarrow{p_{\text{mer}}} \{(11,(i,j,k)_1)\},$$

где $d,l \in \{1,\ldots,6\}$, причём $d \neq l$.

Заметим, что номера кластеров в этом случае заданы согласно маркировке, проведённой на предыдущей итерации.

3) Оккупируются свободные клетки при помощи оператора $\theta_1(i,j,k)$.
4) Формируется новая кластерная структура из оккупированных клеток при помощи маркировки кластеров вспомогательным клеточным автоматом №2.

Таким образом, каждая итерация завершается уничтожением кластерной структуры, образованной на предыдущем шаге, и формированием новой кластерной структуры с автоматическим обновлением всех характеристик кластеров. Конечной стадией изменения во времени (эволюции) кластерной структуры считается конфигурация, в которой образуется кластер, соединяющий противоположные грани куба. Заметим, что приведённый анализ работы автомата позволяет отнести его к асинхронному типу.

3. Результаты модельного эксперимента

Моделирование проводилось на кубической решётке $100 \times 100 \times 100$. При этом для однородного статического сценария моделирования постоянные вероятности имели значения $p_{\text{occ}} = p_{\text{spr}} = p_{\text{mer}} = 0{,}001$, а для внутреннего динамического сценария $p_{\text{occ}} = 0{,}0001$, начальное значение вероятности прорастания периметра $p_{\text{spr}} = 0{,}18$, $p_{\text{mer}} = 0{,}2$. Полученные характеристики случайных процессов усреднялись по 10 реализациям. Пример визуализации соединяющего кластера приведён на рис. 1; верхняя грань соединена с нижней и левой гранями.

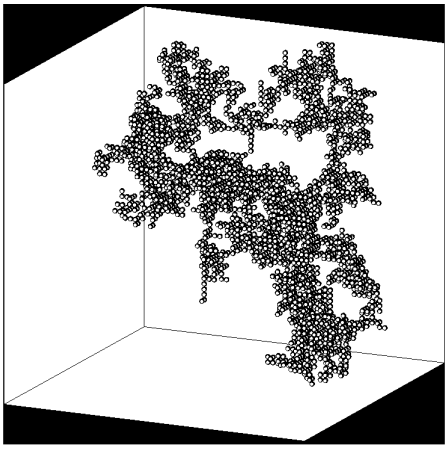


Рис. 1. Вид соединяющего кластера повреждений для динамического внутреннего сценария моделирования

Конфигурация кластеров повреждений имеет фрактальную структуру, о чём свидетельствует универсальная степенная зависимость (рис. 2) между числом элементарных повреждений в кластере (массой кластера) и его среднеквадратичным радиусом $M(R) = 3^3 \cdot R^D$, где $2{,}305 < D < 2{,}336$.

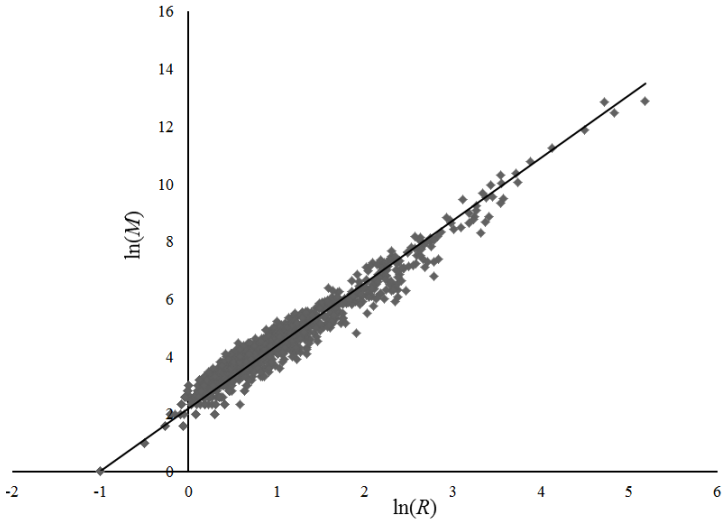


Рис. 2. Связь массы и размера кластеров

Предельная средняя плотность элементарных повреждений, при которой происходит образование соединяющего кластера, составила $d_{\text{fin}} = 0{,}27 \pm 0{,}02$ для однородного статического сценария и $d_{\text{fin}} = 0{,}08 \pm 0{,}02$ для динамического внутреннего сценария, что в 1,2–3 раза меньше порога перколяции на кубической решётке, составляющего $d_{\text{fin}} = 0{,}311$ [12].

На рис. 3 представлено сопоставление кинетических кривых числа кластеров элементарных повреждений для двумерного и трёхмерного случаев моделирования при сопоставимых параметрах моделирования. Для сравнения временных рядов числа кластеров при различных сценариях данные представлены в нормированных координатах: отношение числа кластеров к максимальному числу кластеров и отношение числа циклов к числу циклов до образования соединяющего кластера. Как видно из рисунка, накопление числа кластеров на начальной стадии процесса разрушения в случае трёхмерной модели происходит более медленно как для однородного статического, так и для динамического внутреннего сценариев моделирования. Кроме того, для трёхмерной модели отсутствует протяжённый линейный участок на временах $T/T_{\text{fin}} \geqslant 0{,}55$, характерный для двумерного случая [10]. График кинетической зависимости числа кластеров надёжно аппроксимируется по методу наименьших квадратов квадратичной функцией вплоть до появления соединяющего кластера и разрушения системы: $N/N_{\max} = -2{,}79\,(T/T_{\text{fin}})^2 + 3{,}36(T/T_{\text{fin}})$ с коэффициентом детерминации $R^2 = 0{,}99$ для однородного статического сценария и $N/N_{\max} = -1{,}91\,(T/T_{\text{fin}})^2 + 2{,}80(T/T_{\text{fin}})$ с коэффициентом детерминации $R^2 = 0{,}99$ для внутреннего динамического сценария. Максимумы числа кластеров для трёхмерного случая смещены в сторону больших значений и достигаются при $T/T_{\text{fin}} \approx 0{,}73$.

Как видно из рис. 4,*б*, 5,*б* и 6,*б*, по характеру изменения автокорреляционных функций во временном ряду «число кластеров» имеют место долговременные корреляции, которые соответствуют виду кинетических кривых и имеют универсальный колебательный характер.

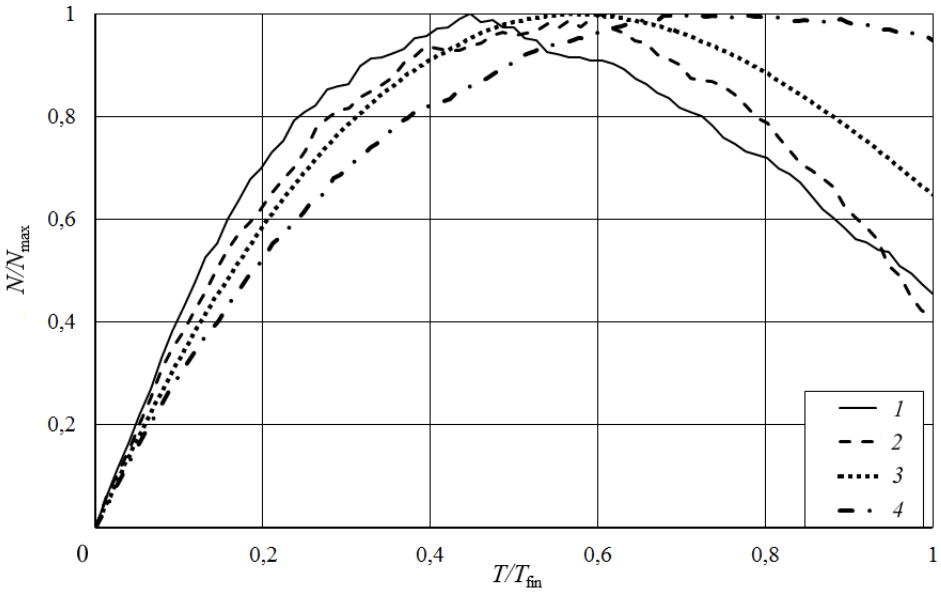


Рис. 3. Кинетическая зависимость числа кластеров элементарных повреждений: кр. *1* — однородный статический сценарий (2D); кр. *2* — динамический внутренний сценарий (2D); кр. *3* — однородный статический сценарий (3D); кр. *4* — динамический внутренний сценарий (3D)

Поведение временного ряда «число элементарных повреждений» существенно зависит от выбора параметров моделирования. Для динамического внутреннего сценария кинетика накопления элементарных повреждений в трёхмерной модели в зависимости от значения вероятности прорастания периметра кластеров повреждений p_{spr} показывает два качественно различных режима моделирования. Так, для случая $p_{\text{spr}} < 0{,}2$ ($p_{\text{spr}} = 0{,}18$) временной ряд «число элементарных повреждений», обнаруживая рост на первых шагах эволюции, флуктуирует вблизи практически не подверженного тренду среднего значения. Число кластеров повреждений значительно превосходит число вновь возникающих элементарных повреждений (рис. $4,а$). Во временном ряду «число элементарных повреждений» имеют место долговременные корреляции (рис. $4,б$) с выходом корреляционной функции в отрицательную область. При значениях $p_{\text{spr}} = 0{,}2$ число вновь возникающих элементарных повреждений и число кластеров повреждений совпадают по порядку величины (рис. $5,а$). Процесс образования соединяющего кластера ускоряется примерно в 3 раза. Процессы, формирующие временной ряд «число элементарных повреждений», становятся более совпадающими как между собой, так и с процессом формирования кластеров повреждений, что проявляется в поведении корреляционных функций (рис. $5,б$). Для значений вероятности прорастания периметров кластеров $p_{\text{spr}} > 0{,}2$ ($p_{\text{spr}} = 0{,}22$) процесс объединения элементарных повреждений и возникновения соединяющего кластера настолько ускоряется, что число вновь возникающих элементарных повреждений в несколько раз превышает число кластеров элементарных повреждений (рис. $6,а$), и соответственно число итераций до появления соединяющего кластера существенно сокращается. Процессы, формирующие временной ряд «число элементарных повреждений», и процесс формирования кластеров становятся полностью совпадающими. Корреляционные функции также практически совпадают, что говорит о возникновении сильных долговременных корреляций (рис. $6,б$).

Эффективным тестом для проверки того, является ли изучаемый процесс процессом с независимыми приращениями, персистентным или антиперсистентным, являет-

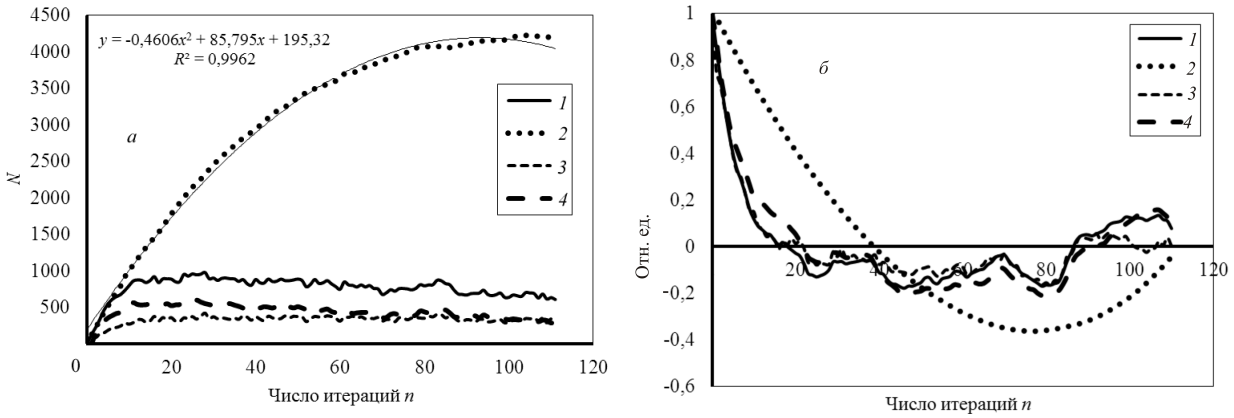


Рис. 4. Кинетические зависимости ($а$) и автокорреляционные функции ($б$) временных рядов «число элементарных повреждений» и «число кластеров повреждений» при $p_{\text{spr}} = 0{,}18$: кр. *1* — суммарная кривая элементарных повреждений; кр. *2* — число кластеров повреждений; кр. *3* — число одиночных элементарных повреждений; кр. *4* — число периметров кластеров

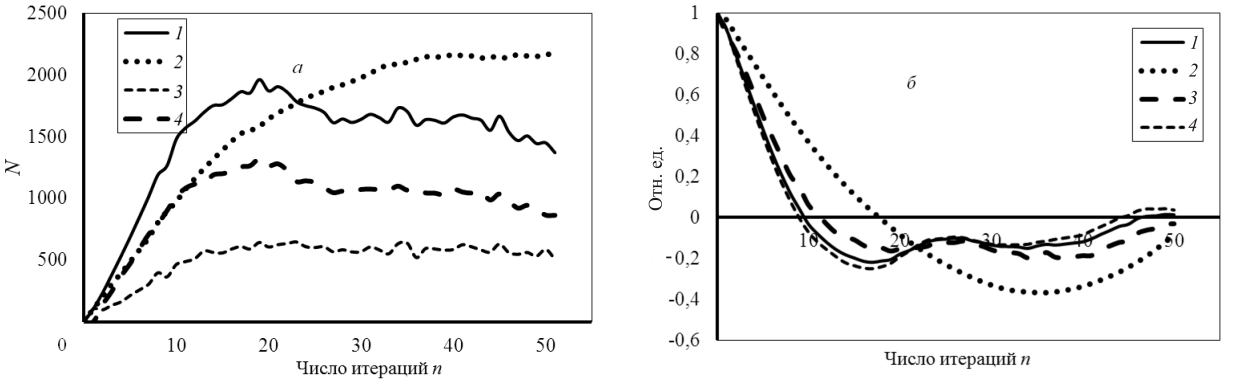


Рис. 5. Кинетические зависимости ($а$) и автокорреляционные функции ($б$) временных рядов «число элементарных повреждений» и «число кластеров повреждений» при $p_{\text{spr}} = 0{,}2$: кр. *1* — суммарная кривая элементарных повреждений; кр. *2* — число кластеров повреждений; кр. *3* — число одиночных элементарных повреждений; кр. *4* — число периметров кластеров

ся метод нормированного размаха Херста (R/S-анализ) [12 – 14]. При этом исходные данные — временные зависимости текущего значения размаха выборки R случайного процесса, нормированного на текущее значение среднеквадратичного отклонения S ($R(t)/S(t) \infty |t|^H$), приводят в дважды логарифмических координатах к линейной зависимости $\ln R/S \infty H \cdot \ln |t|$, угловой коэффициент которой H и называется показателем Херста. Случайному процессу с независимыми приращениями соответствует значение показателя Херста $H = 0{,}5$. Для персистентного случайного процесса (в будущем поддерживается тенденция, которая была в предшествующие моменты времени) значение показателя Херста $H > 0{,}5$. Как и в случае двумерной модели [14], для всех рассмотренных сценариев моделирования временные ряды «число элементарных повреждений» и «число кластеров элементарных повреждений» являются персистентными (рис. 7–9). На рисунках под временем понимаем число итераций n. При этом показатель Херста для случайного процесса «число кластеров элементарных повреждений» составляет $H = 0{,}98 \pm 0{,}01$. Для случайного процесса «число элементарных повре-

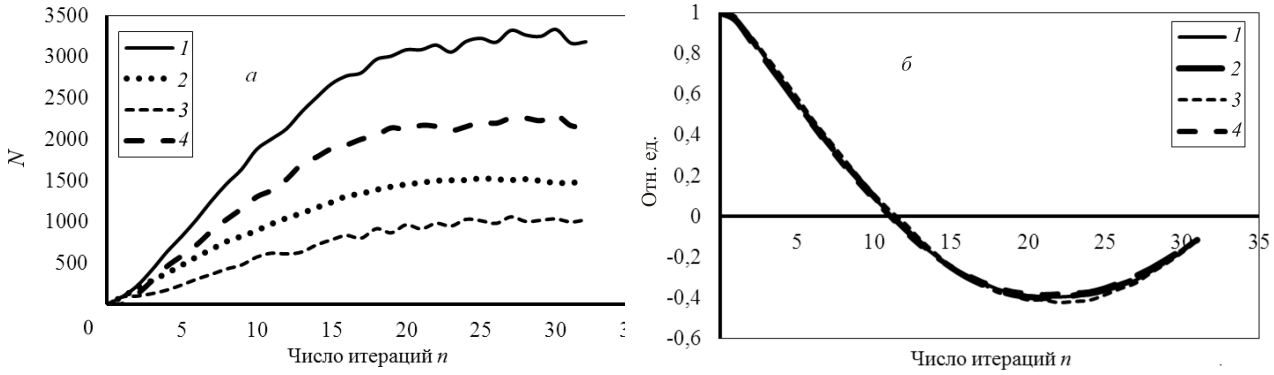


Рис. 6. Кинетические зависимости ($а$) и автокорреляционные функции ($б$) временных рядов «число элементарных повреждений» и «число кластеров повреждений» при $p_{\text{spr}} = 0{,}22$: кр. *1* — суммарная кривая элементарных повреждений; кр. *2* — число кластеров повреждений; кр. *3* — число одиночных элементарных повреждений; кр. *4* — число периметров кластеров

ждений» на статистике нормированного размаха наблюдаются два линейных участка, при этом второй линейный участок, с увеличенным показателем Херста, начинается на временах, составляющих примерно 60–70 % от времени разрушения системы. Исключение составляет случай, когда вероятность прорастания периметра $p_{\text{spr}} > 0{,}2$. Этот режим моделирования приводит к сильной корреляции всех случайных процессов, и, как следствие, все линейные зависимости нормированного размаха совпадают.

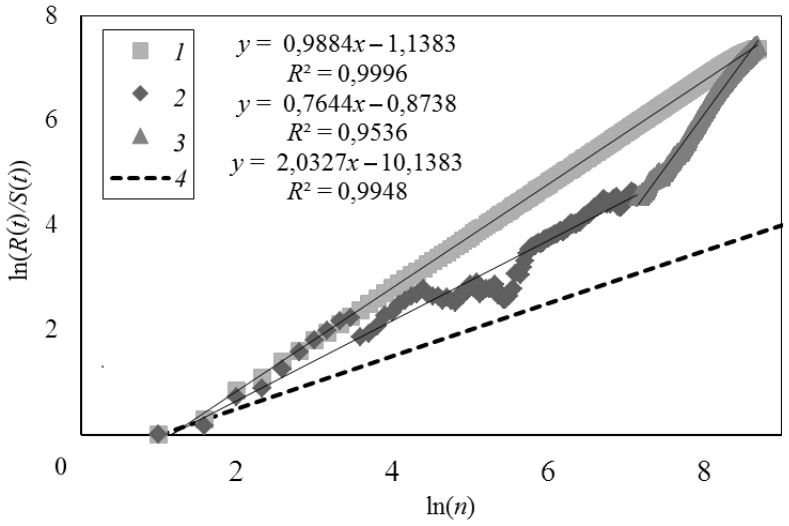


Рис. 7. Зависимость нормированного размаха Херста от числа итераций для базового однородного статического сценария моделирования: кр. *1* — кластеры; кр. *2*, *3* — элементарные повреждения (первый и второй участки); кр. *4* — линия, соответствующая значению $H = 0{,}5$

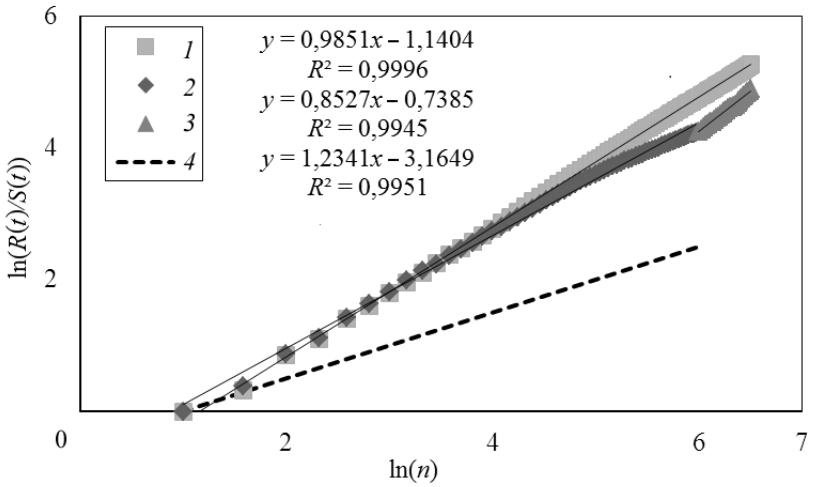


Рис. 8. Зависимость нормированного размаха Херста от числа итераций для динамического внутреннего сценария ($p_{\text{spr}} = 0{,}18$) моделирования: кр. *1* — кластеры; кр. *2*, *3* — элементарные повреждения (первый и второй участки); кр. *4* — линия, соответствующая значению $H = 0{,}5$

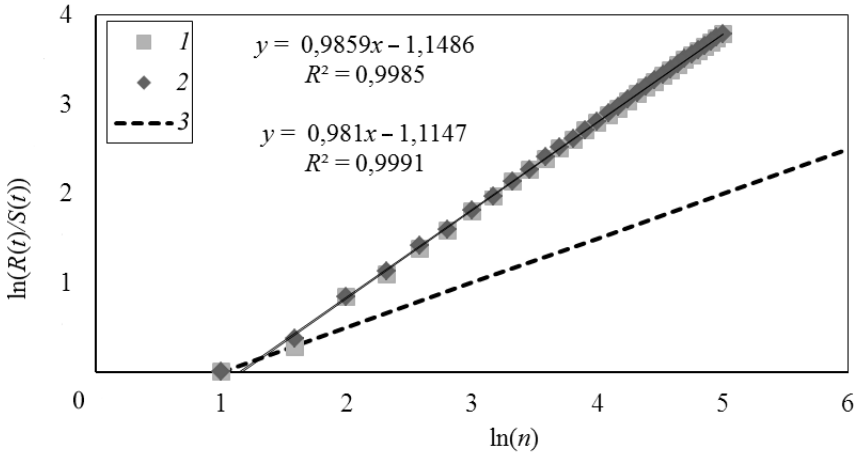


Рис. 9. Зависимость нормированного размаха Херста от числа итераций для динамического внутреннего сценария ($p_{\text{spr}} = 0{,}22$) моделирования: кр. *1* — кластеры; кр. *2* — элементарные повреждения; кр. *3* — линия, соответствующая значению $H = 0{,}5$

4. Сравнение модельного и физического эксперимента

Поскольку наблюдение кластеров элементарных повреждений в динамике на современном уровне технологии практически невозможно, данные модельного эксперимента для временного ряда «число элементарных повреждений» сопоставляем с характеристиками потока импульсной эмиссии. При этом поведение автокорреляционных функций числа импульсов эмиссии (как световой, так и электромагнитной) качественно совпадает с поведением автокорреляционной функции временного ряда «число элементарных повреждений» для режима $p_{\text{spr}} = 0{,}18$ (рис. 10) [15]. С результатами модельного эксперимента также качественно согласуется появление второго линейного

участка на статистике нормированного размаха Херста временного ряда «число импульсов эмиссии» (рис. 11).

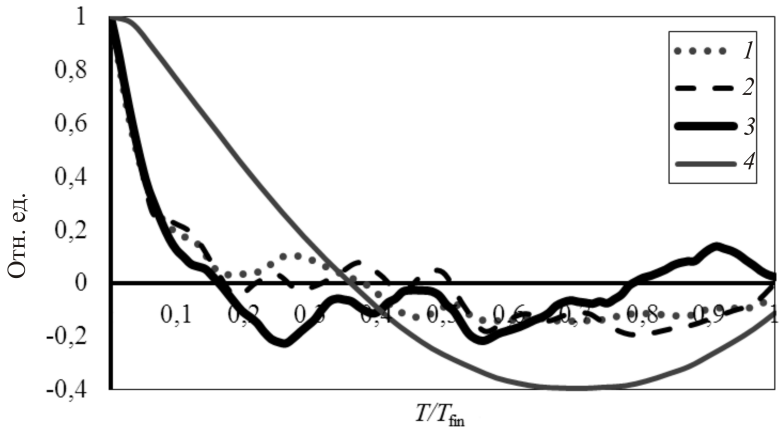


Рис. 10. Автокорреляционные функции временного ряда «число элементарных повреждений»: кр. *1* — кварцевый диорит, электромагнитная эмиссия; кр. *2* — кварцевый диорит, фотонная эмиссия; кр. *3* — внутренний динамический сценарий $p_{\text{spr}} = 0{,}18$; кр. *4* — внутренний динамический сценарий $p_{\text{spr}} = 0{,}22$

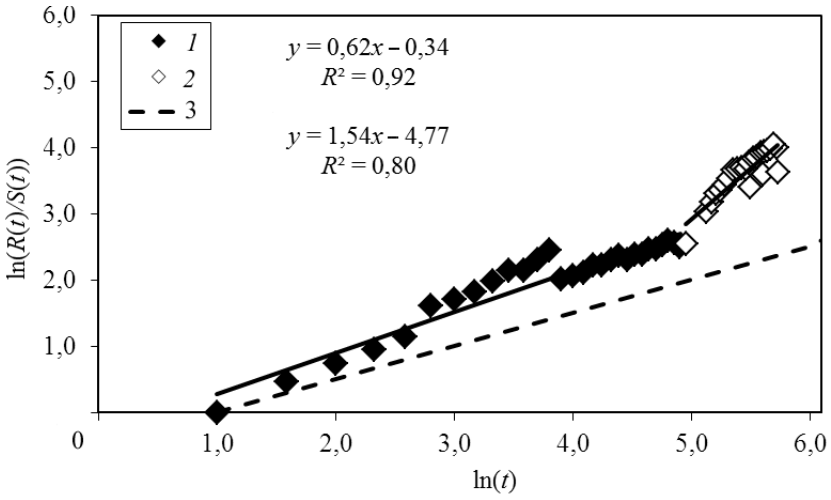


Рис. 11. Статистика нормированного размаха Херста временного ряда «число импульсов электромагнитной эмиссии» для кварцевого диорита: кр. *1* — первый линейный участок; кр. *2* — второй линейный участок ($T/T_{\text{fin}} = 0{,}70$); кр. *3* — линия, соответствующая значению $H = 0{,}5$

Заключение

Построена и реализована трёхмерная модель накопления элементарных повреждений в хрупких гетерогенных материалах при помощи вероятностного клеточного автомата. Установлено, что в зависимости от значения вероятности прорастания периметра кластера повреждений для трёхмерной модели наблюдается два качественно различных режима процесса накопления повреждений. Для значений $p_{\text{spr}} > 0{,}2$ процесс перехода к необратимому разрушению существенно ускоряется и, по сравнению

с режимом при $p_{\text{spr}} < 0{,}2$, становится сильно коррелированным. При этом наблюдается соответствие данных модельного и физического экспериментов: рассчитанное по данным физического эксперимента поведение автокорреляционных функций числа импульсов эмиссии (световой, электромагнитной) и поведение статистики Херста качественно совпадают с поведением автокорреляционной функции временного ряда «число элементарных повреждений» для режима моделирования $p_{\text{spr}} < 0{,}2$. Поэтому наблюдаемый в автокорреляционных функциях потоков импульсов эмиссии переход в область отрицательной корреляции, а также появление второго линейного участка на статистике нормированного размаха можно рассматривать как предвестников перехода системы на стадию необратимого разрушения.

ЛИТЕРАТУРА

1. *Куксенко В. С.* Диагностика и прогнозирование разрушения крупномасштабных объектов // ФТТ. 2005. Т. 47. № 5. С. 788–792.
2. *Курленя М. В., Вострецов А. П., Кулаков Г. И., Яковицкая Г. Е.* Регистрация и обработка сигналов электромагнитного излучения горных пород. Новосибирск: СО РАН, 2000. 232 с.
3. *Ботвина Л. Р.* Разрушение: кинетика, механизмы, общие закономерности. М.: Наука, 2008. 334 с.
4. *Бандман О. Л.* Клеточно-автоматные модели пространственной динамики // Системная информатика. 2006. Вып. 10. С. 59–113.
5. *Бандман О. Л.* Дискретное моделирование физико-химических процессов // Прикладная дискретная математика. 2009. № 3. С. 33–49.
6. *Лобанов А. И.* Моделирование клеточных автоматов // Компьютерные исследования и моделирование. 2010. Т. 2. № 3. С. 273–293.
7. *Гиляров В. Л., Варкентин М. С., Корсуков В. Е. и др.*Формирование степенных распределений дефектов по размерам в процессе разрушения материалов // ФТТ. 2010. Т. 52. № 7. С. 1311–1315.
8. *Гиляров В. Л.* Моделирование роста трещин в процессе разрушения гетерогенных материалов // ФТТ. 2011. Т. 53. № 4. С. 707–710.
9. *Алексеев Д. В., Казунина Г. А.* Моделирование кинетики накопления повреждений вероятностным клеточным автоматом // ФТТ. 2006. Т. 48. № 2. С. 255–261.
10. *Алексеев Д. В., Казунина Г. А.* Моделирование эволюции кластерной структуры элементарных повреждений в нагруженных материалах // Деформация и разрушение материалов. 2009. № 8. С. 10–14.
11. *Гулд Х., Тобочник Я.* Компьютерное моделирование в физике. Ч. 2. М.: Мир, 1990. 390 с.
12. *Федер Е.* Фракталы. М.: Мир, 1991. 258 с.
13. *Алексеев Д. В., Егоров П. В.* Персистентность накопления трещин при нагружении горных пород и концентрационный критерий разрушения // Доклады АН. 1993. № 1. С. 779–780.
14. *Алексеев Д. В., Казунина Г. А.* Модельное исследование кинетики накопления повреждений методом нормированного размаха Херста // Физико-технические проблемы разработки полезных ископаемых. 2006. № 4. С. 69–74.
15. *Казунина Г. А., Малышин А. А.* Исследование кинетики накопления повреждений в нагруженных материалах по импульсной электромагнитной и фотонной эмиссии // Изв. вузов. Физика. 2009. Т. 52. № 6. С. 46–49.

REFERENCES

1. *Kuksenko V. S.* Diagnostika i prognozirovanie razrusheniya krupnomasshtabnykh ob"ektov [Diagnosis and prognosis of large-scale objects destruction]. Fizika Tverdogo Tela, 2005, vol. 47, no. 5, pp. 788–792. (in Russian)
2. *Kurlenya M. V., Vostretsov A. P., Kulakov G. I., and Yakovitskaya G. E.* Registratsiya i obrabotka signalov elektromagnitnogo izlucheniya gornykh porod [Register and Processing of the Signals of the Rocks Electromagnetic Radiation]. Novosibirsk, SB RAS, 2000. 232 p.(in Russian)
3. *Botvina L. R.* Razrushenie: kinetika, mekhanizmy, obshchie zakonomernosti [Destruction: Kinetics, Mechanisms, General Patterns]. Moscow, Nauka Publ., 2008. 334 p. (in Russian)
4. *Bandman O. L.* Kletochno-avtomatnye modeli prostranstvennoy dinamiki [Cellular automata models of spatial dynamics]. Sistemnaya Informatika, 2006, no. 10, pp. 59–113. (in Russian)
5. *Bandman O. L.* Diskretnoe modelirovanie fiziko-khimicheskikh protsessov [Discrete models of physical-chemical processes]. Prikladnaya Diskretnaya Matematika, 2009, no. 3, pp. 33–49. (in Russian)
6. *Lobanov A. I.* Modelirovanie kletochnykh avtomatov [Model of cellular automata]. Komp'yuternye issledovaniya i modelirovanie, 2010, vol. 2, no. 3, pp. 273–293. (in Russian)
7. *Gilyarov V. L., Varkentin M. S., Korsukov V. E., et al.*Formirovanie stepennykh raspredeleniy defektov po razmeram v protsesse razrusheniya materialov [Formation of power-law distributions of defects in size during the fracture of materials]. Fizika Tverdogo Tela, 2010, vol. 52, no. 7, pp. 1311–1315. (in Russian)
8. *Gilyarov V. L.* Modelirovanie rosta treshchin v protsesse razrusheniya geterogennykh materialov [Simulation of crack growth in the destruction of heterogeneous materials]. Fizika Tverdogo Tela, 2011, vol. 53, no. 4, pp. 707–710. (in Russian)
9. *Alekseev D. V. and Kazunina G. A.* Modelirovanie kinetiki nakopleniya povrezhdeniy veroyatnostnym kletochnym avtomatom [Modeling the kinetics of damage accumulation of probabilistic cellular automata]. Fizika Tverdogo Tela, 2006. vol. 48, no. 2, pp. 255–261. (in Russian)
10. *Alekseev D. V. and Kazunina G. A.* Modelirovanie evolyutsii klasternoy struktury elementarnykh povrezhdeniy v nagruzhennykh materialakh [Modeling the evolution of the cluster structure of the elementary damage in loaded materials]. Deformatsiya i Razrushenie Materialov, 2009, no. 8, pp. 10–14. (in Russian)
11. *Guld Kh. and Tobochnik Ya.* Komp'yuternoe modelirovanie v fizike [Computer Modeling in Physics]. P. 2. Moscow, Mir Publ., 1990, 390 p. (in Russian)
12. *Feder J.* Fractals. N.Y., Springer, 1988.
13. *Alekseev D. V. and Egorov P. V.* Persistentnost' nakopleniya treshchin pri nagruzhenii gornykh porod i kontsentratsionnyy kriteriy razrusheniya [Persistence accumulation of cracks during loading of rock and concentration failure criterion]. Doklady Akademii Nauk, 1993, no. 1, pp. 779–780. (in Russian)
14. *Alekseev D. V. and Kazunina G. A.* Model'noe issledovanie kinetiki nakopleniya povrezhdeniy metodom normirovannogo razmakha Khersta [Model study of the damage accumulation kinetics by the Hirst normed span method]. Fiziko-tekhnicheskie problemy razrabotki poleznykh iskopaemykh, 2006, no. 4, pp. 69–74. (in Russian)
15. *Kazunina G. A. and Mal'shin A. A.* Study of the kinetics of damage accumulation in loaded materials based on impulse electromagnetic and photon emission. Russian Physics J., 2009, no. 6, pp. 598–601.

СВЕДЕНИЯ ОБ АВТОРАХ

АЛЕКСЕЕВ Дмитрий Валентинович — доктор технических наук, профессор, профессор Кемеровского института (филиала) Российского экономического университета им. Г. В. Плеханова, г. Кемерово. E-mail: **dmitriyalekseev@live.ru**

АНИСЕНЯ Николай Ильич — студент кафедры защиты информации и криптографии Национального исследовательского Томского государственного университета, г. Томск. E-mail: **anisenya@gmail.com**

БРОСЛАВСКИЙ Олег Викторович — студент кафедры защиты информации и криптографии Национального исследовательского Томского государственного университета, г. Томск. E-mail: **o.v.broslavsky@gmail.com**

ГОРНОВА Маргарита Николаевна — студентка Омского государственного университета им. Ф. М. Достоевского, г. Омск. E-mail: **mgornova@gmail.com**

ДЕНИСОВ Олег Викторович — кандидат физико-математических наук, ООО «Центр сертификационных исследований», г. Москва.
E-mail: **denisovOleg@yandex.ru**

КАЗУНИНА Галина Алексеевна — доктор технических наук, доцент, профессор кафедры математики Кузбасского государственного технического университета им. Т. Ф. Горбачева, г. Кемерово. E-mail: **gt-kga@yandex.ru**

КОЛЕГОВ Денис Николаевич — кандидат технических наук, доцент кафедры защиты информации и криптографии Национального исследовательского Томского государственного университета, г. Томск. E-mail: **d.n.kolegov@gmail.com**

КУКИНА Екатерина Георгиевна — кандидат физико-математических наук, директор Института математики и информационных технологий Омского государственного университета им. Ф. М. Достоевского, г. Омск. E-mail: **katpop@ya.ru**

МИНАКОВ Александр Александрович — преподаватель Московского государственного технического университета радиотехники, электроники и автоматики, г. Москва. E-mail: **minak-ski@yandex.ru**

НИКОЛАЕВ Максим Владимирович — аспирант кафедры информационной безопасности Московского государственного университета им. М. В. Ломоносова, г. Москва, Россия. E-mail: **max.abstract@gmail.com**

ОЛЕКСОВ Никита Евгеньевич — студент кафедры защиты информации и криптографии Национального исследовательского Томского государственного университета, г. Томск. E-mail: **n.e.oleksov@gmail.com**

РОМАНЬКОВ Виталий Анатольевич — доктор физико-математических наук, профессор, заведующий кафедрой информационных систем Омского государственного университета им. Ф. М. Достоевского, г. Омск. E-mail: **romankov48@mail.ru**

РЫБАЛОВ Александр Николаевич — кандидат физико-математических наук, доцент кафедры математической логики и логического программирования Омского государственного университета им. Ф. М. Достоевского, г. Омск.
E-mail: **alexander.rybalov@gmail.com**

ФОМИЧЕВ Владимир Михайлович — доктор физико-математических наук, профессор, профессор Финансового университета при Правительстве Российской Федерации, заместитель по науке технического директора ООО «Код безопасности», г. Москва. E-mail: **fomichev@nm.ru**

ФРОЛОВА Юлия Юрьевна — кандидат физико-математических наук, доцент кафедры алгебро-геометрических вычислений Ульяновского государственного университета, г. Ульяновск. E-mail: **yuyufrolova@mail.ru**

ЧЕРЕДНИЧЕНКО Алла Валериевна — аспирантка кафедры математики Кузбасского государственного технического университета им. Т. Ф. Горбачева, г. Кемерово. E-mail: **allacherednichenk@rambler.ru**

ШУЛЕЖКО Олеся Владимировна — аспирантка Ульяновского государственного университета, г. Ульяновск. E-mail: **ol.shulezhko@gmail.com**

ШУРУПОВ Андрей Николаевич — кандидат технических наук, доцент Московского государственного технического университета радиотехники, электроники и автоматики (МИРЭА), г. Москва. E-mail: **ashurupov@mail.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, 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*