Некоторые свойства бесповторных слов

Л.М. Волков, Е.В. Суханов

Рассматриваются слова в алфавите из двух или трех символов. Эти объекты изучаются с точки зрения свойств повторяемости блоков. Находится точная мера бесповторности для широкого класса бескубных слов. Устанавливается максимальная длина бесквадратного слова, не содержащего наперед заданного блока.

1. Введение

В начале века А. Туэ открыл и впервые исследовал так называемые бесквадратные и бескубные последовательности в алфавитах из двух и трех букв. Позднее были найдены и другие аналогичные последовательности. Такие объекты широко используются в комбинаторике и до сих пор их свойства привлекают внимание исследователей (см. по этому поводу, например, недавние работы [2] и [4]).

Предлагаемая работа является продолжением предыдущих изысканий на эту тему. Мы указываем в ней некоторые новые свойства таких (как говорят, бесповторных) последовательностей. Постановки задач и полученные результаты содержатся в разделах 2 и 3. Основные определения и обозначения вынесены в данный раздел.

Напомним, прежде всего, что *алфавитом* называется произвольное конечное непустое множество Σ, а его элементы называют *буквами*. В наших рассмотрениях будут встречаться только алфавиты из двух или трех букв: $\Sigma_2 = \{a, b\}$ и $\Sigma_3 = \{a, b, c\}$. *Слово над* Σ – это конечная цепочка букв этого алфавита. Пустое слово обозначается через λ. Число символов в записи слова W называется его длиной и обозначается через $|W|$. Мы будем рассматривать и бесконечные слова, то есть последовательности вида $W = x_1x_2\ldots$, где $x_i \in \Sigma$. Тот факт, что в i-той позиции слова W стоит буква x, мы будем обозначать равенством $W(i) = x$. Конечное слово V называется *подсловом* конечного или бесконечного слова W, если V состоит из таких n букв, что в W найдется позиция i со свойством $V(1) = W(i)$, $V(2) = W(i+1)$, …, $V(n) = W(i+n-1)$. Отношение «быть подсловом» частично упорядочивает множество всех слов над данным алфавитом, поэтому мы иногда будем прибегать к записи $V \leqslant W$. Запись $W[i,j]$ (с непременным условием $i \leqslant j$) будет использоваться для обозначения подслова в W, которое начинается в i-той и заканчивается на j-той позиции. Слово λ является подсловом любого слова. На множестве конечных слов определена ассоциативная операция *конкатенации* (приписывания): XY обозначает слово, состоящее из слова X и написанного сразу за ним слова Y. Очевидны равенства $X\lambda = \lambda X = X$, $|XY| = |X| + |Y|$ и $|X^n| = n \cdot |X|$.

Пусть в непустом слове W для некоторого числа p и для всех номеров позиций $i \leqslant |W|$ выполняется равенство $W(i) = W(i+p)$, если $i+p \leqslant |W|$. Тогда число p называется *периодом* слова W. Любое слово обладает хотя бы одним периодом, так как для произвольного числа $p \geqslant |W|$ верно, что $i+p > |W|$, и поэтому p – это период. Наименьший период слова W мы будем обозначать через $\mathrm{per}(W)$. Например, $\mathrm{per}(abaac) = 5$, $\mathrm{per}(aabbaabb) = 4$, а $\mathrm{per}(abcab) = 3$.

Экспонентой слова W называется число $\exp(W) = |W|/\mathrm{per}(W)$. Так, например, $\exp(abcab) = 5/3$, $\exp(abac) = 1$, а $\exp(aaa) = 3$. Говорят, что конечное или бесконечное непустое слово W *избегает* экспоненту γ, если для любого его подслова V выполняется неравенство $\exp(V) < \gamma$. В противном случае говорят, что W *допускает* экспоненту γ.

Множество $\Gamma(W) = \{\gamma : W$ избегает экспоненту $\gamma\}$ всегда ограничено снизу единицей и представляет собой луч на множестве всех действительных чисел, так как если W избегает экспоненту γ, то, по определению, W избегает и экспоненту $\gamma+\varepsilon$ для любого $\varepsilon > 0$. *Пределом избегаемости* непустого слова W (конечного или бесконечного)

называется число $\mathrm{limexp}(W) = \inf \Gamma(W)$. Справедливо, конечно, и такое равенство: $\mathrm{limexp}(W) = \sup\{\gamma : W$ допускает экспоненту $\gamma\}$.

Предел избегаемости – это ключевое понятие для данной работы, так как главной проблемой, решаемой в ней, является вычисление предела избегаемости некоторых важных и часто встречающихся бесконечных слов.

Конечное или бесконечное слово W называется *бесквадратным* (*бескубным*), если оно не содержит подслов вида X^2 (соответственно X^3), где X – произвольное непустое слово. На языке избегаемости бесквадратность означает, что слово избегает экспоненту 2, а бескубность – что слово избегает экспоненту 3. Слово W называется *сильно бескубным*, если оно не содержит подслов вида $xYxYx$, где x – это буква, а Y – слово (возможно пустое). Легко понять, что сильно бескубное слово избегает экспоненту $2+\varepsilon$ для любого $\varepsilon > 0$.

Для примера займемся сначала вычислением предела избегаемости двух важных бесконечных слов, которые принято называть словами Туэ-Морса. Первое из них – мы будем обозначать его через TM_2 – является сильно бескубным над двухбуквенным алфавитом Σ_2. Чтобы построить TM_2, рассмотрим последовательность $\{W_n\}$, которая удовлетворяет следующим свойствам: $W_1 = a$ и $W_i = W_{i-1}\overline{W_{i-1}}$, где $\overline{W_{i-1}}$ есть слово, полученное из W_{i-1} заменой всех вхождений буквы a на b и наоборот. Тогда начало последовательности $\{W_n\}$ будет выглядеть следующим образом: a, ab, $abba$, $abbabaab\ldots$.

Последовательность $\{W_n\}$ называют последовательностью Туэ-Морса над двухбуквенным алфавитом. Через TM_2 мы обозначаем такое бесконечное слово, что любое его начальное подслово длины 2^n совпадает с W_{n+1}, то есть $TM_2 = abbabaabba\ldots$. Известно (см., например, [3]), что любое слово последовательности $\{W_n\}$ и все слово TM_2 являются сильно бескубными. Отсюда немедленно вытекает, что $\mathrm{limexp}(TM_2) = 2$, так как с одной стороны TM_2 допускает экспоненту 2 (ведь, например, $bb \leqslant TM_2$), а с другой стороны, в силу своей сильной бескубности, TM_2 избегает экспоненту $2+\varepsilon$ для любого $\varepsilon > 0$.

Построим теперь бесконечное бесквадратное слово TM_3 над алфавитом Σ_3. Пусть $X = TM_2[i, i+1]$. Тогда i-тая буква слова TM_3 определяется по правилу

$$TM_3(i) = \begin{cases} a, & \text{если} \quad X = ab \\ b, & \text{если} \quad X = ba \\ c, & \text{если} \quad X = aa \quad \text{или} \quad X = bb. \end{cases}$$

Известно (см., например, [3]), что полученное таким образом бесконечное слово $TM_3 = acbabc\ldots$ – слово Туэ-Морса над трехбуквенным алфавитом – является бесквадратным. Найдем теперь предел избегаемости для TM_3.

Предложение 1. *Предел избегаемости слова TM_3 равен 2, причем это слово избегает экспоненту 2.*

Доказательство. Рассмотрим слово K_n, состоящее из первых 2^n букв слова TM_3. Тогда слово, состоящее из первых 2^{n+2} букв TM_3, можно будет вычислить по формуле $K_{n+2} = K_n\overline{K_n}\overline{K_n}K_n$. Отсюда немедленно следует, что подслова $V_1 = TM_3[2^n+1, 2^{n+1}-1]$ и $V_2 = TM_3[2^{n+1}+1, 2^{n+1}+2^n-1]$ совпадают и $|V_1| = |V_2| = 2^n - 1$. Рассмотрим теперь $V = TM_3[2^n+1, 2^{n+1}+2^n-1]$. Его длина равна $2^{n+1}-1$. Из равенства $V_1 = V_2$ вытекает, что V обладает периодом величины 2^n. Поэтому $\mathrm{per}(V) \leqslant 2^n$ и тогда

$$\exp(V) = \frac{|V|}{\mathrm{per}(V)} \geqslant \frac{2^{n+1}-1}{2^n}.$$

Проведенные выше вычисления проходят для любого натурального n. Следовательно,

$$\mathrm{limexp}(TM_3) \geqslant \lim_{n\to\infty} \frac{2^{n+1}-1}{2^n} = 2.$$

С другой стороны, в силу бесквадратности TM_3, мы имеем неравенство $\mathrm{limexp}(TM_3) \leqslant 2$. А это значит, что нам удалось вычислить предел избегаемости слова Туэ-Морса над трехбуквенным алфавитом – он равен 2, причем TM_3 экспоненту 2 избегает.

2. Гомоморфные образы бесквадратных слов

В 1937 году в [1] С.И. Аршоном была построена последовательность бесквадратных слов над Σ_3. От этой последовательности мы будем отталкиваться в дальнейших рассмотрениях раздела, поэтому сначала мы воспроизведем построение Аршона.

Рассмотрим перестановки букв трехбуквенного алфавита. Назовем ***нечетными*** перестановки abc, bca и cab. Они занумерованы своими первыми буквами. Перестановки cba, acb и bac будем называть четными. Они занумерованы своими последними буквами.

Положим $U_1 = a$. Пусть теперь $U_i = x_1x_2\ldots x_n$. Тогда, чтобы получить слово U_{i+1}, необходимо заменить каждую букву x_j четной или нечетной перестановкой (в зависимости от четности числа j), начинающейся или, соответственно, заканчивающейся буквой x_j. Так, $U_2 = abc$, а $U_3 = abcacbcab$. В своей работе Аршон показал, что все слова U_i бесквадратны.

Слово Аршона над Σ_3 – это бесконечное слово A_3, каждое начало которого длины 3^i совпадает с U_{i+1}. В [2] доказано, что A_3 не только бесквадратно, но выполняется и более сильное равенство $\mathrm{limexp}(A_3) = 7/4$.

Аршоном был рассмотрен гомоморфизм ν, действующий из полугруппы слов над алфавитом Σ_3 в полугруппу слов над Σ_2 и определенный на образующих элементах следующим образом:

$$\nu(a) = ab; \qquad \nu(b) = aba; \qquad \nu(c) = abba.$$

Под гомоморфным образом бесконечного слова мы будем понимать результат замены каждой буквы x_i на слово $\nu(x_i)$, то есть $\nu(W) = \nu(x_1x_2\ldots) = \nu(x_1)\nu(x_2)\ldots$. Аршон показал, что слово $A_2 = \nu(A_3)$ является бескубным. Этот результат можно обобщить и усилить. Ниже (см. теорему 1) мы вычислим предел избегаемости широкого класса слов, который содержит и A_2.

Пусть бесконечное слово B над Σ_3 является бесквадратным. Тогда *обобщенным словом Аршона* над Σ_2 мы будем называть слово

$A = \nu(B)$. Чтобы вычислить предел избегаемости обобщенных слов Аршона, нам понадобятся два вспомогательных утверждения.

Лемма 1. *Пусть*

1) $U = ab$,

2) V *– слово над* Σ_2,

3) W, X *– слова над* Σ_3.

Если $UV = \nu(X)$ *и* $UVUVU \leqslant \nu(W)$, *то* $XX \leqslant W$.

Доказательство. Пусть Y и Z – слова над Σ_3. Покажем, что если $\nu(Y) \geqslant \nu(Z)U$, то $Y \geqslant Z$. В самом деле, $\nu(Z)U = UV_1UV_2\ldots V_nU$, где $n = |Z|$, а $V_i \in \{\lambda, a, ba\}$. Расшифровка этого слова однозначно определяется последовательностью $V_1, \ldots, V_n$ (так как каждое слово U, не неся в себе никакой специальной информации, отмечает только начало образа новой буквы слова Z), каждое из которых задает одну букву слова Z. Поэтому если $\nu(Y)$ содержит $\nu(Z)U$, то Y содержит X.

По условию леммы слово $UVUVU = \nu(X)\nu(X)U = \nu(XX)U$ является подсловом в $\nu(W)$. Следовательно, $XX \leqslant W$.

Лемма 2. *Любое бесконечное бесквадратное слово над алфавитом* Σ_3 *содержит подслово* cba.

Доказательство. Бесконечное бесквадратное слово U над Σ_3 содержит букву c, так как над алфавитом Σ_2 длина максимального бесквадратного слова равна 3. Покажем теперь, что U содержит подслово cb. В противном случае каждый раз после буквы c следовала бы буква a. Ниже рассмотрено слово $V \leqslant U$, начинающееся с буквы c и не содержащее cb. В левом столбце таблицы производится постепенное наращивание V. Заглавные буквы – это те, которые добавляются на очередном шаге наращивания. В правом столбце рассмотрены другие возможности (кроме самых очевидных), и показано, как они приводят к противоречию.

$caB\ldots$	–	нельзя $ca\,CA$ (из-за бесквадратности)
$cabA\ldots$	–	нельзя $cab\,CAB$ или $cab\,CA\,CA$
$cabaC\ldots$	–	конечно нельзя $c\,ab\,aB$
$caba\,cABA$	–	получаем квадрат

Табл. 1

В таблице показано, что каждая следующая буква слова V определяется единственным образом, но даже и такое построение приводит к противоречию с бесквадратностью V. Поэтому бесконечного бесквадратного слова, не содержащего подслова cb, не существует. Покажем теперь, что U содержит и подслово cba. Попробуем построить бесконечное бесквадратное слово, в котором после подслова cb каждый раз идет буква c. Результат построения отображен в таблице:

$cbcAB\ldots$	–	нельзя $cbcA\,CBCA$
$cbcabAC\ldots$	–	нельзя $cbca\,bC\,BC$; $c\,bca\,bCA$
$cbcabacAB\ldots$	–	нельзя $cbcaba\,cBCABA$
$cbcabacabCA\ldots$	–	$cb\,caba\,cabA$; $cbcabaca\,bC\,BC$
$cbcabacabcaCB\ldots$	–	$cbcaba\,cab\,caB$
$cbcabacabca\,cbCABACABCA$	–	получаем квадрат

Табл. 2

Мы показали, что бесконечное бесквадратное слово не может не содержать подслова cba – лемма доказана.

Теорема 1. *Предел избегаемости любого обобщенного слова Аршона над двухбуквенным алфавитом равен 8/3, причем экспоненту 8/3 такое слово допускает.*

Доказательство. Пусть A – некоторое обобщенное слово Аршона. Пусть также V некоторое подслово в A. Наша цель – установить, какой максимальной длины может достигать слово V при заданном периоде $\mathrm{per}(V)$. Разберем возникающие здесь возможности.

I. $\mathrm{per}(V) = 1$. Легко понять, что в этом случае $|V| \leqslant 2$. В самом деле, если $|V| = 3$, то (так как наименьший период V равен 1) либо $V = aaa$, либо $V = bbb$. Однако такие слова V не могут являться подсловами обобщенного слова Аршона A – это следует из структуры гомоморфизма ν. Итак, для тех слов V, у которых наименьший период равен 1, выполняется неравенство $\exp(V) = |V|/\mathrm{per}(V) \leqslant 2$.

II. $\mathrm{per}(V) = 2$. В этом случае $|V| \leqslant 5$. Для $|V| = 6$ есть снова две возможности: $V_1 = ababab$ и $V_2 = bababa$. Попробуем найти прообраз слов V_1 и V_2 в B. Напомним, что $A = \nu(B)$ для подходящего бесквадратного слова B. В нечетных строках таблицы 3 записаны V_1 и V_2, а в четных (крупным шрифтом) – те буквы слова B, которые однозначно восстанавливаются из условия $A = \nu(B)$.

для слова V_1	$A:$		ab	ab	ab
	$B:$		a	a	
для слова V_2	$A:$	b	ab	ab	a
	$B:$	a	a		

Табл. 3

Таким образом, в обоих случаях мы получаем противоречие с бесквадратностью слова B (у нас получилось, что $aa \leqslant B$). Отсюда следует, что при $\mathrm{per}(V) = 2$ верно неравенство $\exp(V) \leqslant 5/2$.

III. $\mathrm{per}(V) = 3$. Покажем, что в этом случае $|V| \leqslant 8$. Для доказательства используем тот же прием, что и в предыдущем случае – попробуем найти прообразы для слов V с периодом 3 длины 9 (знак $\emptyset$ в таблице 4 означает, что прообраза не существует).

Эта таблица наглядно показывает, что для слов $V \leqslant A$ длины 9 либо не существует прообраза в B, либо этот прообраз содержит подслово bb. Значит, в случае $\mathrm{per}(V) = 3$ можно утверждать: $\exp(V) \leqslant 8/3$. Заметим также, что $\nu(cbab) = ab\,baa\,baa\,ba\,ba$, а $\nu(cbac) = ab\,baa\,baa\,ba\,bba$. По лемме 2 любое бесконечное бесквадратное слово B содержит подслово cba (а, следовательно, и одно из подслов $cbab$ или $cbac$). Поэтому любое обобщенное слово Аршона A содержит подслово $baabaaba$ и допускает экспоненту $8/3$.

для $V = bbabbabba$	$A:$	bb	abba	bba	
	$B:$	$\emptyset$	c	$\emptyset$	
для $V = babbabbab$	$A:$	b	abba	bb	ab
	$B:$	a	c	$\emptyset$	
для $V = abbabbabb$	$A:$	abba	bb	abb	
	$B:$	c	$\emptyset$	$\emptyset$	
для $V = aabaabaab$	$A:$	a	aba	aba	ab
	$B:$		b	b	
для $V = abaabaaba$	$A:$	aba	aba	ab	a
	$B:$	b	b		
для $V = baabaabaa$	$A:$	ba	aba	aba	a
	$B:$		b	b	

Табл. 4

IV. $\mathrm{per}(V) \geqslant 4$. Мы докажем, что в этом случае $|V| \leqslant 2\cdot\mathrm{per}(V)+2$. Для этого достаточно установить следующий факт. Если $V = XXY$, где Y – некоторое начало слова X, а $|X| = \mathrm{per}(V)$, то $|Y| \leqslant 2$.

Выделим в X первое вхождение подслова ab. Это вхождение может начинаться с первой, второй, третьей или четвертой позиции слова X. Рассмотрим эти случаи, начиная с последнего.

1) Четвертая позиция. Очевидно, что если слово ab появляется в X впервые только начиная с четвертой позиции, то ему предшествует подслово bba – окончание $\nu(c)$. Тогда слово V имеет вид

$$\overbrace{\underbrace{a\,bbaab\ldots}_{X}\underbrace{a\,bbaab\ldots}_{X}\underbrace{a\,bbaab\ldots}_{Y}}^{V}$$

Чтобы свести этот случай к первому, достаточно только сдвинуть все фигурные скобки на одну позицию влево.

2) Третья позиция. Рассмотрим ситуацию, когда X имеет вид $x_1x_2ab\ldots$. Тогда слово x_1x_2 является окончанием одного из слов

$\nu(b)$ или $\nu(c)$, то есть $X = baab\ldots$. Предположим теперь, что слово Y – это начало X длины не менее чем 3. В этом случае $Y = baa\ldots$, а значит $Y = baab\ldots$, так как (это было показано выше) aaa не может являться подсловом A_2. Итак, мы пришли к такой картине:

$$\underbrace{ba\ \overbrace{ab}^{U}\ \overbrace{x_5\ldots}^{V}}_{X}\ \underbrace{ba\ \overbrace{ab}^{U}\ \overbrace{x_5\ldots}^{V}}_{X}\ \underbrace{ba\ \overbrace{ab\ y_5\ldots}^{U}}_{Y}$$

Применяя к $UVUVU$ лемму 1, мы немедленно приходим к противоречию с бесквадратностью исходного слова B. Таким образом, длина Y в этом случае не может превышать 2.

3) Вторая или первая позиция. Аналогичное рассуждение доказывает, что в случае $X = x_1ab\ldots$ длина слова Y также не больше 2, а в случае $X = ab\ldots$ (к которому нам удалось свести и случай $X = bbaab\ldots$) длина Y не больше 1. Проиллюстрируем эти случаи рисунками:

$$\underbrace{x_1\ \overbrace{ab}^{U}\ \overbrace{x_4\ldots}^{V}}_{X}\ \underbrace{x_1\ \overbrace{ab}^{U}\ \overbrace{x_4\ldots}^{V}}_{X}\ \underbrace{x_1\ \overbrace{ab\ y_4\ldots}^{U}}_{Y}$$

$$\underbrace{\overbrace{ab}^{U}\ \overbrace{x_3\ldots}^{V}}_{X}\ \underbrace{\overbrace{ab}^{U}\ \overbrace{x_3\ldots}^{V}}_{X}\ \underbrace{\overbrace{ab\ y_3\ldots}^{U}}_{Y}$$

Итак, нам удалось показать, что если $\mathrm{per}(V) \geqslant 4$, то

$$\exp(V) = \frac{|V|}{\mathrm{per}(V)} = \frac{2\cdot|X| + |Y|}{\mathrm{per}(V)} \leqslant \frac{2\cdot \mathrm{per}(V) + 2}{\mathrm{per}(V)} = 2 + \frac{2}{\mathrm{per}(V)} \leqslant \frac{5}{2}$$

Этим доказательство теоремы завершается, так как теперь мы знаем, что при любом значении $\mathrm{per}(V)$, где V – произвольное подслово обобщенного слова Аршона A, выполнено неравенство $\exp(V) \leqslant 8/3$, причем равенство достигается (при $V = baabaaba$).

Примененный выше способ доказательства можно применять также и для других гомоморфизмов; нужно только, чтобы они удовлетворяли условиям леммы 1. Рассмотрим, например, гомоморфизм ξ, который определен на буквах алфавита Σ_3 следующим образом:

$$\xi(a) = a, \quad \xi(b) = ab, \quad \xi(c) = abb.$$

Легко проверить, что лемма 1 справедлива для гомоморфизма ξ. Теперь аналогично теореме 1 нетрудно доказать такое утверждение.

Теорема 2. *Пусть B – произвольное бесквадратное слово над алфавитом Σ_3 и $A = \xi(B)$. Тогда* $\mathrm{limexp}(A) \leqslant 5/2$.

Мы не случайно привели здесь в качестве примера гомоморфизм ξ. Он понадобится нам и в следующем параграфе, где его использование будет играть решающую роль. Пока заметим лишь, что ξ инъективен на своем образе. Поэтому можно говорить об обратном отображении ξ^{-1}, определенном на множестве слов над Σ_2, не содержащих bbb.

3. Слова, не содержащиеся в бесквадратных

Если некоторое бесконечное слово W над алфавитом Σ_3 бесквадратно, то оно не содержит aa, bb, cc, $abab$ и прочих слов вида XX в качестве своих подслов. Однако понятно, что бесквадратные слова могут не содержать и других подслов. Так, например, ни одно бесконечное бесквадратное слово не содержит $abacaba$ – слова Зимина над трехбуквенным алфавитом (о словах Зимина см., например, [7]), так как, хотя это слово и бесквадратно, какую из букв a, b или c мы бы не добавили, все равно получим квадрат.

С другой стороны, выше было показано (лемма 2), что любое бесконечное бесквадратное слово обязано содержать подслово ab и даже abc. Таким образом, естественно возникают следующие вопросы:

1) *Пусть задано некоторое бесквадратное слово V. Существует ли бесконечное бесквадратное слово W, которое не содержит подслова V?*

2) Если такое слово есть, то привести пример; если нет, то какой максимальной длины может достигать бесквадратное слово, не содержащее V?

На эти вопросы мы дадим ниже исчерпывающий ответ. Здесь уместно отметить, что аналогичные вопросы для алфавита из четырех букв и экспоненты 3/2 рассматривались в работе [5].

Теорема 3. *Пусть X и W – бесквадратные слова над Σ_3, причем W не содержит подслова, совпадающего с X. Тогда*

1) Если $|X| = 1$, то $|W| \leqslant 3$,

2) Если $|X| = 2$, то $|W| \leqslant 13$,

3) Если $|X| = 3$ и все буквы слова X различны, то $|W| \leqslant 30$,

4) Для любого другого X существует бесконечное слово W, которое его не содержит.

Доказательство.

1) Первое утверждение, конечно, вполне очевидно. Пусть, например, $X = c$. Тогда бесквадратное слово W длины 3, не содержащее X – это либо aba, либо bab. Легко понять, что четырехбуквенных слов с такими свойствами не существует.

2) В лемме 2 было показано, что любое бесконечное бесквадратное слово обязано содержать подслово cb. Естественно, что совершенно безразлично, каким из двухбуквенных бесквадратных слов является слово X, так как, переобозначив буквы, мы всегда сможем прийти к случаю $X = cb$. Из доказательства этой леммы также следует, что самое длинное слово, которое начинается с букв ca, бесквадратно и не содержит подслова cb – это $V = cabacabcac$. Буква c (по пункту 1) обязана встретиться самое позднее на четвертой позиции в искомом слове W максимальной длины. Поэтому $W = babV$ (где $|W| = 13$) – самое длинное бесквадратное слово, не содержащее cb.

3) Это утверждение также следует из леммы 2. Самое длинное слово над трехбуквенным алфавитом, которое не содержит cba – это $W = bacabacbcabacabcacbcabacbcabcb$. Его длина равна 30. Мы опускаем несложные, но объемные вычисления, с помощью которых этот результат выводится из доказательства леммы 2. Наметим лишь основной путь. Рассмотрим слово cbc и будем наращивать его справа, пользуясь построением, проведенным в вышеупомянутой лемме. Каждый раз, получив cbc, мы будем вынуждены пойти по другому, более короткому пути, чтобы не получить противоречия с бесквадратностью. Когда же дальнейшее удлинение вправо станет невозможным, начнем так же продолжать W влево. Можно показать, что это делается единственным способом. Скоро мы и здесь исчерпаем все возможности, получив слово W. В самом деле, в силу симметричности cbc, наращивание его в обе стороны происходит абсолютно одинаковым образом, однако наращивание влево прервется раньше. Дело в том, что после ab мы не можем добавить слева c, так как читая W слева направо мы получим в нем подслово cba.

4) Будем говорить, что конечные слова X_1 и X_2 над алфавитом Σ *имеют один вид*, если существует такая перестановка φ букв Σ, что $X_2 = \varphi(X_1)$.

Например, слова $abac$ и $bcba$ – одного вида (мы также будем называть их словами вида $xyxz$), а слова aba и bac – разного.

Кроме слов вида xyz, которые были рассмотрены выше, бывают еще трехбуквенные бесквадратные слова вида xyx. Бесконечное бесквадратное слово W, которое мы сейчас построим, не будет содержать подслова cbc. Следовательно, какое бы мы ни взяли слово X вида xyx, из W соответствующей перестановкой букв удастся получить слово, которое не содержит X. Более того, мы покажем, что среди четырехбуквенных слов, которые не содержат W, есть слова всех видов. Значит, существует бесконечное бесквадратное слово, которое не содержит любого наперед заданного четырехбуквенного бесквадратного слова.

Рассмотрим бесконечное слово $TM_2 = abbabaab\ldots$ и гомоморфизм ξ – эти объекты уже упоминались выше. Однозначно определяется слово $W = \xi^{-1}(TM_2)$. В самом деле, расшифровка производится следующим образом: находим очередное вхождение буквы a и считаем, сколько букв b встретятся в TM_2 до следующего ее вхождения. Из бескубности слова TM_2 следует, что букв b будет не более двух, и нам удастся корректно определить очередную букву слова W.

Известно (см. [6]), что полученное таким образом W совпадает со словом Туэ-Морса над трехбуквенным алфавитом TM_3 с точностью до переобозначения букв. В дальнейшем мы, не обращая внимания на тот факт, что при каноническом построении слова TM_3 оно отличается от слова W наименованем букв, будем обозначать слово $W = \xi^{-1}(TM_2)$ через TM_3.

Как уже отмечалось выше, слово $W = TM_3$ является бесквадратным. Исследуем теперь вопрос о вхождении в TM_3 подслов интересующих нас видов, используя равенство $TM_2 = \xi(TM_3)$. Очевидно, TM_3 не содержит подслова cbc, равно как и подслова aba. В самом деле, $\xi(cbc) = ab\ babab\ b$, но слово $babab$ не является сильно бескубным и поэтому не может содержаться в TM_2. Если бы TM_3 содержало подслово aba не в начале, то оно содержало бы и подслово $cabac$. Но $\xi(cabac) = ab\ baabaab\ b$, а подслово $baabaab$ также не является сильно бескубным. Легко убедиться, что слово aba не может быть и началом TM_3.

Обратимся теперь к четырехбуквенным бесквадратным словам. Есть три вида таких слов: $xyxz$, $zxyx$ и $xyzx$. Из того, что TM_3 не содержит подслова cbc, немедленно следует, что оно не содержит и подслов $cbca$ и $acbc$, которые являются словами соответственно первого и второго видов. Легко понять также, что TM_3 не содержит подслов $bcab$ и $bacb$, которые принадлежат к третьему виду. В самом деле, вместе с подсловом $bcab$ слово TM_3 содержало бы и подслово $abcabc$ (так как перед bc обязательно стоит буква a, а после ab – буква c), что противоречит его бесквадратности. Совершенно аналогичное рассуждение можно применить и к слову $bacb$.

Таким образом, к какому виду четырехбуквенных слов не принадлежало бы X, в бесконечном слове TM_3 можно так переобозначить буквы, чтобы оно не содержало подслова X. Понятно теперь, что это верно и для любых бесквадратных слов X с числом букв больше 4, так как любое такое слово содержит четырехбуквенное бесквадратное подслово.

Уместно отметить, что существует несколько способов получения бесконечных бесквадратных слов, не содержащих заданного подслова, например cbc. Однако здесь мы рассматривали именно слово $W = TM_3$, поскольку из всех известных авторам примеров оно является единственным, которое позволяет снять все рассматривавшиеся выше вопросы, касающиеся трех- и четырехбуквенных бесквадратных подслов.

Список литературы

[1] Аршон С.Е. Доказательство существования n-значных бесконечных асимметричных последовательностей // Мат. сб. 1937. Т. 2(44). №4. С. 769–779.

[2] Клепинин А.В., Суханов Е.В. О комбинаторных свойствах языка Аршона // Дискретный анализ и исследование операций. 1999. №2.

[3] Саломаа А. Жемчужины теории формальных языков. М.: Наука, 1985.

[4] Шур А.М. Бинарная избегаемость и слова Туэ-Морса // Доклады РАН. 1996. Т. 348. №5. С. 598–599.

[5] Brandenburg F.-J. Uniformly growing k-th power-free homomorphisms // Theoretical Computer Science. 1983. Vol. 23. P. 69–82.

[6] Lothaire M. Combinatorics on words // Encyclopedia of Mathematics and its Applications. Vol. 17. Addison-Wesley, Reading, MA, 1983. Reprinted in the «Cambridge Mathematical Library». Cambridge: Cambridge University Press, 1997.

[7] Sapir M.V. Combinatorics on words with applications // Institute Blaise Pascal, Rapports LITP, 1995. Vol. 32.