G. V. Balakin, On the possibility of partial recovery of some sequences via observations, *Mat. Vopr. Kriptogr.*, 2013, Volume 4, Issue 4, 7–25

DOI: 10.4213/mvk97

МАТЕМАТИЧЕСКИЕ ВОПРОСЫ КРИПТОГРАФИИ
2013 Т. 4 № 4 С. 7–25

УДК 519.212.2

О возможности частичного восстановления некоторых последовательностей по наблюдениям

Г. В. Балакин

Академия криптографии Российской Федерации, Москва

Получено 20.V.2011

В работе предлагаются методы частичного восстановления некоторых последовательностей по известным наблюдениям над их знаками или s-граммами, $s \geq 2$.

Ключевые слова: системы нелинейных уравнений, рекуррентные последовательности, искаженные наблюдения

On the possibility of partial recovery of some sequences via observations

G. V. Balakin

Academy of Cryptography of the Russian Federation, Moscow

Abstract. We describe methods of partial recovery of some sequences by means of known observations on their elements or s-gramms, $s \geq 2$.

Key words: systems of nonlinear equations, recurrent sequences, noisy observations

Citation: *Mathematical Aspects of Cryptography*, 2013, vol. 4, no. 4, pp. 7–25 (Russian).

1. Постановка задачи

Пусть имеется последовательность

$$x_1, x_2, \ldots, x_t \tag{1}$$

символов из множества $\{0,1,\ldots,q-1\}$, обладающая некоторыми структурными особенностями. Эти особенности нам частично известны; дополнительно могут быть известны наблюдения над отдельными членами последовательности (1) и (или) s-граммами $x_i, x_{i+1}, \ldots, x_{i+s-1}$ этой последовательности, $s \geq 2$. Ставится задача по известной информации уточнить структурные особенности последовательности и восстановить некоторые участки неизвестной последовательности (1), либо обнаружить, что задача не имеет решения.

Такая задача возникает, например:

- при восстановлении марковской последовательности с известными запрещенными переходами по наблюдениям над s-граммами,
- при восстановлении потока данных, состоящего из серий одинаковых знаков, по наблюдениям над знаками,
- при рассмотрении систем уравнений сдвигового типа,
- при нахождении трехчленной рекурренты по наблюдениям над знаками,
- при решении систем уравнений рекуррентного типа.

В п. 2 изучается система уравнений сдвигового типа с равновероятной пороговой функцией в левой части. Число булевых неизвестных $2m+1$ в функции неизвестно. Предлагается метод определения возможных значений числа m и частичного определения знаков исходной последовательности. Применение метода продемонстрировано на двух примерах.

Частный случай системы уравнений рекуррентного типа подробно рассмотрен в п. 3. В левой части системы находится пороговая функция от трех неизвестных — линейных форм от членов трехчленной рекурренты. Приведены методы нахождения значений отдельных линейных форм и сведения системы нелинейных уравнений к системе линейных псевдобулевых уравнений, зависящих от двух или трех неизвестных.

В п. 4 предлагаются методы нахождения параметров трехчленной рекурренты над полем $GF(q)$ по наблюдениям над ее членами.

В п. 5 предполагается, что элементы последовательности (1) принадлежат множеству $\{0,1,\ldots,q-1\}$, $q \geq 3$. Над этими элементами проводятся наблюдения. Приводится метод частичного восстановления исходной последовательности.

2. Система уравнений сдвигового типа

Пусть булева последовательность (1) удовлетворяет системе уравнений

$$F\left(x_i, x_{i+1}, \ldots, x_{i+2m}\right) = a_i, \quad i = 1, 2, \ldots, t-2m. \tag{2}$$

Функция F пороговая, так что

$$x_i + x_{i+1} + \ldots + x_{i+2m} \geq m+1 \text{ при } a_i = 1,$$

$$x_i + x_{i+1} + \ldots + x_{i+2m} \leq m \text{ при } a_i = 0.$$

Число m может быть неизвестным.

Лемма 1. *Пусть для некоторого* $i \in \{1, 2, \ldots, t-2m-1\}$ *выполняется равенство* $a_i + a_{i+1} = 1$. *Тогда*

$$x_i = a_i, \; x_{i+2m+1} = a_i \oplus 1 = a_{i+1}.$$

Доказательство. Действительно, если $a_i = 1$, $a_{i+1} = 0$, то

$$x_i + x_{i+1} + \ldots + x_{i+2m} \geq m+1,$$

$$x_{i+1} + x_{i+2} + \ldots + x_{i+2m+1} \leq m.$$

Из этих неравенств получаем неравенства

$$m - x_{i+2m+1} \geq x_{i+1} + x_{i+2} + \ldots + x_{i+2m} \geq m+1-x_i,$$

из которых следует, что $x_i = 1$, $x_{i+2m+1} = 0$. Второй случай $a_i = 0$, $a_{i+1} = 1$ рассматривается аналогично.

Следствие 1. *Если* $a_i + a_{i+1} = 1$, *то*

$$x_{i+1} + x_{i+2} + \ldots + x_{i+2m} = m.$$

Следствие 2. *Если для некоторого* m *найдется такое* $i \in \{1, 2, \ldots, t-2m-2\}$, *что*

$$a_i = 1, \; a_{i+1} = 0, \; a_{i+2m+1} = 1, \; a_{i+2m+2} = 0$$

либо

$$a_i = 0, \; a_{i+1} = 1, \; a_{i+2m+1} = 0, \; a_{i+2m+2} = 1,$$

то система уравнений (2) *не имеет решений.*

Доказательство. Пусть

$$x_i + x_{i+1} + \ldots + x_{i+2m} \geq m+1,$$

$$x_{i+1} + x_{i+2} + \ldots + x_{i+2m+1} \leq m.$$

Тогда согласно лемме 1 и следствию 1

$$x_i = 1, x_{i+2m+1} = 0, x_{i+1} + x_{i+2} + \ldots + x_{i+2m} = m .$$

Далее, из неравенств

$$x_{i+2m+1} + x_{i+1+2m+2} + \ldots + x_{i+4m+1} \geq m+1 ,$$

$$x_{i+2m+2} + x_{i+1+2m+3} + \ldots + x_{i+4m+2} \leq m$$

находим

$$x_{i+2m+1} = 1, x_{i+4m+2} = 0, x_{i+2m+2} + x_{i+2m+3} + \ldots + x_{i+4m+1} = m-1 .$$

Таким образом, мы приходим к противоречию в определении значения x_{i+2m+1}. Аналогично рассматривается второй случай.

Метод нахождения значений отдельных знаков из последовательности (1) при известном значении m заключается в следующем. Вначале мы находим первую знакоперемену в последовательности (1), пусть это будет $a_i = 0$, $a_{i+1} = 1$. Определяем значения первого и последнего знаков из тех знаков, которые присутствуют в этих двух уравнениях. Эти значения подставляем в уравнения (2) и переходим ко второй знакоперемене и так далее. В результате этой процедуры мы определим ряд неизвестных и найдем псевдобулевы линейные ограничения (равенства и неравенства) для остальных неизвестных. В частности, на первом этапе мы определим, что $x_i = 0$, $x_{i+2m+1} = 1, x_i + x_{i+1} + \ldots + x_{i+2m} = m,\quad x_{i+1} + x_{i+2} + \ldots + x_{i+2m+1} = m+1$, а для $j = 1, 2, \ldots, i-1$ верны неравенства

$$x_j + x_{j+1} + \ldots + x_{j+2m} \leq m .$$

Трудозатраты этой процедуры пропорциональны t.

Таким образом, если последовательность (1) соответствует совместной системе уравнений (2), то она имеет ряд структурных особенностей. В частности, согласно лемме 1 в ней отсутствуют одинаковые знакоперемены на расстоянии $2m+1$. Следовательно, если параметр m не является известным, то мы можем, используя вышеприведенное свойство, найти возможные значения для m. Подтверждением правильности выбора m будет являться нахождение хотя бы одного решения исходной системы (2).

Рассмотрим пример. Пусть $m = 2, n = 21, t = 25$.

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	
x_i	1	1	1	0	0	1	0	0	1	1	0	0	1	1	1	0	0	0	1	0	0	1	0	1	1
a_i	1	1	0	0	0	1	0	0	1	1	1	1	1	0	0	0	0	0	0	0	1				

Последовательность $\overline{x} = x_1, x_2, \ldots, x_{25}$ взята из таблицы случайных чисел (столбец 2, строка 12), приведенной в [1], в ней имеется 12 нулей и 12 знакоперемен. В последовательности $\overline{a} = a_1, a_2, \ldots, a_{21}$ всего 6 знакоперемен. Используя эти знакоперемены и доказательство леммы 1, находим одиннадцать неизвестных

$x_2 = 1, x_5 = 0, x_6 = 1, x_7 = 0, x_8 = 0, x_{10} = 1, x_{11} = 0, x_{13} = 1, x_{18} = 0, x_{20} = 0, x_{25} = 1,$

неизвестное $x_{13} = 1$ найдено два раза.

Далее из первых трех неравенств $x_1 + x_3 + x_4 \geq 2$, $x_3 + x_4 \geq 1$, $x_3 + x_4 \leq 1$ определяем $x_1 = 1, x_3 + x_4 = 1$. Значения $x_9 = 1$ и $x_{12} = 0$ находятся соответственно из 6-го и 8-го неравенств, значения $x_{14} = x_{15} = 1$ и $x_{16} = x_{17} = 0$ являются решениями 11-го, 12-го, 13-го и 14-го неравенств. Неизвестное x_{19} может принимать любое значение. Из 20-го и 21-го неравенств находим, что $x_{21} + x_{22} + x_{23} + x_{24} = 2$. Следовательно, всего будет 24 решения.

Метод частичного восстановления структурной последовательности, использующий линейные псевдобулевые неравенства, применим и к другим системам. Рассмотрим пример.

Пусть пороговая функция F задается следующим образом:

$$x_i + x_{i+1} - x_{i+2} - x_{i+3} + x_{i+4} \geq 1, \text{ если } F(x_i, x_{i+1}, x_{i+2}, x_{i+3}, x_{i+4}) = 1,$$

$$x_i + x_{i+1} - x_{i+2} - x_{i+3} + x_{i+4} \leq 0, \text{ если } F(x_i, x_{i+1}, x_{i+2}, x_{i+3}, x_{i+4}) = 0.$$

Из двух неравенств

$$x_i + x_{i+1} - x_{i+2} - x_{i+3} + x_{i+4} \geq 1,$$

$$x_{i+1} + x_{i+2} - x_{i+3} - x_{i+4} + x_{i+5} \leq 0$$

находим простые следствия

$$x_{i+1} - x_{i+3} \geq 1 - x_i + x_{i+2} - x_{i+4},$$

$$x_{i+1} - x_{i+3} \leq x_{i+4} - x_{i+5} - x_{i+2},$$

$$1 - x_i \leq 2(x_{i+4} - x_{i+2}) - x_{i+5}.$$

Последовательно определяем значения шести неизвестных:

1. $x_{i+4} = x_{i+2}$, $x_{i+5} = 0$, $x_i = 1$, $x_{i+1} = x_{i+3}$, всего 4 решения;
2. $x_{i+4} = 0$, $x_{i+2} = 1$, третье следствие не имеет решений;
3. $x_{i+4} = 1$, $x_{i+2} = 0$, $-x_i \leq x_{i+1} - x_{i+3} \leq 1 - x_{i+5}$, всего 12 решений.

Итак, вектор из шести неизвестных имеет 16 значений. Это согласуется со схемой, когда одно уравнение имеет 2^4 решений, два последовательно расположенных уравнения совместно имеют тоже 2^4 решений и так далее.

3. Система уравнений рекуррентного типа

Рассмотрим систему уравнений

$$F\left(X\left(\overline{x}_i\right),Y\left(\overline{y}_i\right),Z\left(\overline{z}_i\right)\right)=a_i,i=1,2,\ldots,t\,. \tag{3}$$

Здесь три рекуррентных последовательности

$$\overline{x}=x_1,x_2,\ldots,$$

$$\overline{y}=y_1,y_2,\ldots,$$

$$\overline{z}=z_1,z_2,\ldots,$$

удовлетворяющих одному и тому же рекуррентному соотношению

$$u_i\oplus u_{i+k}\oplus u_{i+n}=0\,,\ i=1,2,\ldots. \tag{4}$$

Вначале рассмотрим случай, когда все три линейные формы X,Y,Z известны, ненулевые и однородные, $\overline{x}_i=\left(x_i,x_{i+1},\ldots,x_{i+n-1}\right)$, аналогично определяются $\overline{y}_i,\overline{z}_i$.

Функция F — пороговая:

$$X\left(\overline{x}_i\right)+Y\left(\overline{y}_i\right)+Z\left(\overline{z}_i\right)\geq 2\text{, если } a_i=1,$$

$$X\left(\overline{x}_i\right)+Y\left(\overline{y}_i\right)+Z\left(\overline{z}_i\right)\leq 1\text{, если } a_i=0\,.$$

Для удобства можно ввести новые обозначения

$$X_i=X\left(\overline{x}_i\right),\ Y_i=Y\left(\overline{y}_i\right),\ Z_i=Z\left(\overline{z}_i\right)$$

и вместо системы (3) рассматривать систему псевдобулевых неравенств относительно сумм $X_i+Y_i+Z_i,i=1,2,\ldots,t$.

Так как булева функция F — пороговая, то систему уравнений (3) можно представить в следующем виде:

$$F\left(X_i,Y_i,Z_i\right)=X_iY_i\oplus X_iZ_i\oplus Y_iZ_i=X_i\oplus\left(X_i\oplus Y_i\right)\left(X_i\oplus Z_i\right)=a_i\,,\ i=1,2,\ldots,t\,.$$

Введем новые булевы неизвестные

$$X_i=a_i\oplus\varepsilon_i\,,$$

где

$$\varepsilon_i=\left(X_i\oplus Y_i\right)\left(X_i\oplus Z_i\right)=L_iQ_i\,,$$

$$L_i = X_i \oplus Y_i\,,\ Q_i = X_i \oplus Z_i\,,\ i = 1,2,\ldots,t\,.$$

Если априори

$$\mathrm{P}\left(X_i = 1\right) = \mathrm{P}\left(Y_i = 1\right) = \mathrm{P}\left(Z_i = 1\right) = 1/2$$

и X_i, Y_i, Z_i как случайные величины независимы, то

$$\mathrm{P}\left(X_i = a_i \,\middle|\, F\left(X_i, Y_i, Z_i\right) = a_i\right) = 3/4\,,\ \mathrm{P}\left(\varepsilon_i = 1\right) = 1 - \mathrm{P}\left(\varepsilon_i = 0\right) = 1/4\,,\ i = 1,\ldots,t\,.$$

Далее заметим, что из булева линейного соотношения

$$L\left(X_{i_1}, X_{i_2}, \ldots, X_{i_s}\right) = b$$

следует, что

$$L\left(\varepsilon_{i_1}, \varepsilon_{i_2}, \ldots, \varepsilon_{i_s}\right) = b \oplus L\left(a_{i_1}, a_{i_2}, \ldots, a_{i_s}\right).$$

Таким образом, если

$$x_i \oplus x_{i+k} \oplus x_{i+n} = 0\,, \tag{5}$$

то

$$\varepsilon_i \oplus \varepsilon_{i+k} \oplus \varepsilon_{i+n} = a_i \oplus a_{i+k} \oplus a_{i+n}\,, \quad i = 1,\ldots,t-n\,. \tag{6}$$

Аналогично можно составить другие соотношения, используя (6). Помимо соотношения (5) можно использовать другие соотношения

$$x_i \oplus x_{i+2^s k} \oplus x_{i+2^s n} = 0, \quad s \geq 1\,.$$

Если X_i есть линейная форма X от неизвестных $x_i, x_{i+1}, \ldots, x_{i+n-1}$, удовлетворяющих равенству (3), то мы приходим к линейной системе уравнений с искаженной правой частью

$$X_i = X\left(x_i, \ldots, x_{i+n-1}\right) = a_i \oplus \varepsilon_i\,, \quad i = 1,\ldots,t\,.$$

При этом ошибки ε_i, $i = 1,\ldots,t$, удовлетворяют линейным булевым уравнениям (6), а неизвестные $x_1, x_2, \ldots$ — соотношению (5).

По существу мы рассматриваем простейшую систему булевых уравнений с искаженной правой частью при ограничениях на значения неизвестных и ошибок [2]. Если $\varepsilon_i = 1$, то будем говорить, что произошло искажение.

Дополнительно можно ввести ошибки и для Y_i, Z_i следующим образом:

$$Y_i = a_i \oplus \xi_i\,,\ Z_i = a_i \oplus \eta_i\,.$$

Тогда мы получим три системы из линейных булевых уравнений с искаженными правыми частями. Как нетрудно видеть, ошибки в правых частях этих уравнений связаны простым псевдобулевым неравенством

$$\varepsilon_i + \xi_i + \eta_i \le 1, \quad i = 1,\ldots,t. \tag{7}$$

Действительно, для $a_i = 0$ это очевидно, а для $a_i = 1$ получим неравенство

$$(1-\varepsilon_i)+(1-\xi_i)+(1-\eta_i) = 3-(\varepsilon_i+\xi_i+\eta_i) \ge 2.$$

Все три линейных формы из системы (3) удовлетворяют одному порождающему соотношению (4). Рассмотрим три уравнения относительно ошибок

$$\varepsilon_i \oplus \varepsilon_{i+k} \oplus \varepsilon_{i+n} = a_i \oplus a_{i+k} \oplus a_{i+n},$$
$$\xi_i \oplus \xi_{i+k} \oplus \xi_{i+n} = a_i \oplus a_{i+k} \oplus a_{i+n},$$
$$\eta_i \oplus \eta_{i+k} \oplus \eta_{i+n} = a_i \oplus a_{i+k} \oplus a_{i+n}.$$

Если правая часть в этих уравнениях равна 1, то, учитывая неравенства (7), находим

$$\varepsilon_i + \varepsilon_{i+k} + \varepsilon_{i+n} = \xi_i + \xi_{i+k} + \xi_{i+n} = \eta_i + \eta_{i+k} + \eta_{i+n} = 1.$$

Число решений этих трех псевдобулевых линейных уравнений относительно девяти ошибок (с учетом (7)) равно 6. Так как в исходной системе (3) с пороговой функцией F есть симметрия относительно трех линейных форм, то для определенности можно положить

$$\varepsilon_i = 1, \varepsilon_{i+k} = \varepsilon_{i+n} = 0, \xi_{i+k} = 1, \xi_i = \xi_{i+n} = 0, \eta_i = \eta_{i+k} = 0, \eta_{i+n} = 1. \tag{8}$$

Дополнительно отметим, что аналогичным свойством обладают и четырехчлены, либо порождающие исходную последовательность $\overline{x} = x_1,\ldots,x_t$, либо являющиеся суммой двух трехчленных соотношений (6) с одним общим членом. Действительно, пусть

$$\varepsilon_i \oplus \varepsilon_{i+k} \oplus \varepsilon_{i+n} = 1, \varepsilon_{i+k} \oplus \varepsilon_{i+2k} \oplus \varepsilon_{i+n+k} = 0.$$

Тогда

$$\varepsilon_i \oplus \varepsilon_{i+2k} \oplus \varepsilon_{i+n} \oplus \varepsilon_{i+n+k} = 1$$

и аналогично, как это сделано для трехчленов, находим

$$\varepsilon_i + \varepsilon_{i+2k} + \varepsilon_{i+n} + \varepsilon_{i+n+k} = 1,$$
$$\xi_i + \xi_{i+2k} + \xi_{i+n} + \xi_{i+n+k} = 1,$$
$$\eta_i + \eta_{i+2k} + \eta_{i+n} + \eta_{i+n+k} = 1.$$

Число решений этих трех уравнений относительно 12 ошибок равно $4 \cdot 3 \cdot 2 = 24$.

Если все три линейные формы в системе (3) зависят от одних переменных $x_1,\ldots,x_n$, то для нахождения решения системы (3) достаточно найти значения n введенных ошибок, находящихся в независимых уравнениях, например, ошибки $\varepsilon_{k+1},\varepsilon_{k+2},\ldots,\varepsilon_{k+n}$. Для $t = 2n$ выбор этих ошибок объясняется тем, что все они входят в два трехчленных равенства вида (6). Остальные ошибки присутствуют только в одном равенстве из системы (6).

Отметим некоторые факты, которые потребуются нам в дальнейшем при поиске решений системы (3).

Введем обозначения: S_i есть i-е уравнение системы (6), которое является суммой трех уравнений с номерами $i, i+k, i+n$ исходной системы (3), $V_i = a_i \oplus a_{i+k} \oplus a_{i+n}$ назовем весом уравнения S_i, $i = 1,2,\ldots,t-n$. Соответственно, уравнение D_i есть сумма из трех уравнений с номерами $i, i+2k, i+2n$ системы (3), $W_i = a_i \oplus a_{i+2k} \oplus a_{i+2n}$ — вес уравнения D_i. Иногда для удобства будем пользоваться обозначением $a(i,j,l) = a_i \oplus a_j \oplus a_l$.

Веса V_i и W_i связаны линейными булевыми соотношениями, например

$$W_i = V_i \oplus V_{i+k} \oplus V_{i+n}.$$

Действительно, рассмотрим три булевых уравнения системы (6)

$$\begin{aligned} \varepsilon_i \oplus \varepsilon_{i+k} \oplus \varepsilon_{i+n} &= V_i, \\ \varepsilon_{i+k} \oplus \varepsilon_{i+2k} \oplus \varepsilon_{i+n+k} &= V_{i+k}, \\ \varepsilon_{i+n} \oplus \varepsilon_{i+n+k} \oplus \varepsilon_{i+2n} &= V_{i+n}. \end{aligned} \tag{9}$$

Сложив эти уравнения, получим

$$W_i = a_i \oplus a_{i+2k} \oplus a_{i+2n} = \varepsilon_i \oplus \varepsilon_{i+2k} \oplus \varepsilon_{i+2n} = V_i \oplus V_{i+k} \oplus V_{i+n},$$

что доказывает утверждение.

Лемма 2. *Для системы уравнений* (6) *множество решений системы из двух булевых уравнений*

$$\begin{aligned} \varepsilon_i \oplus \varepsilon_{i+k} \oplus \varepsilon_{i+n} &= 1, \\ \varepsilon_{i+k} \oplus \varepsilon_{i+2k} \oplus \varepsilon_{i+n+k} &= 0 \end{aligned} \tag{10}$$

совпадает с множеством решений системы из двух псевдобулевых уравнений относительно 5 *неизвестных ошибок*

$$\varepsilon_i + \varepsilon_{i+k} + \varepsilon_{i+n} = 1,$$

$$\varepsilon_{i+k} - \varepsilon_{i+2k} - \varepsilon_{i+n+k} = 0,$$

состоящим из 4 *решений.*

Доказательство. Первое уравнение следует из леммы 3 работы [3]. Второе уравнение по этой лемме заменяется четырехчленным псевдобулевым уравнением

$$\varepsilon_i + \varepsilon_{i+2k} + \varepsilon_{i+n} + \varepsilon_{i+n+k} = 1,$$

из которого следует неравенство

$$\varepsilon_{i+2k} + \varepsilon_{i+n+k} \le 1.$$

Поэтому, если $\varepsilon_{i+k} = 1$, то $\varepsilon_i = \varepsilon_{i+n} = 0$ и $\varepsilon_{i+2k} + \varepsilon_{i+n+k} = 1$, а для $\varepsilon_{i+k} = 0$ находим $\varepsilon_i + \varepsilon_{i+n} = 1, \varepsilon_{i+2k} = 0, \varepsilon_{i+n+k} = 0$. Следовательно, рассматриваемая псевдобулева система из двух уравнений с 5 ошибками имеет 4 решения, и второе булево уравнение можно заменить на псевдобулево равенство.

Если мы хотим получить для случая $X_i = x_i$ линейные псевдобулевы уравнения относительно исходных неизвестных — членов рекурренты, то заменяем ε_j на $x_j \oplus a_j$ в приведенных двух линейных псевдобулевых уравнениях. Для $a_i = 0, a_{i+k} = 1, a_{i+2k} = 1, a_{i+n} = 0, a_{i+n+k} = 1$ получаем два линейных псевдобулевых уравнения

$$x_i - x_{i+k} + x_{i+n} = 0,$$

$$x_{i+2k} + x_{i+n+k} - x_{i+k} = 1,$$

имеющих 4 решения относительно 5 неизвестных членов исходной рекурренты.

Определение 1. Два уравнения назовем *смежными*, если у них имеется одно общее неизвестное.

Определение 2. Три уравнения назовем *смежными*, если каждая из образованных ими трех пар уравнений является смежной и не существует общего неизвестного для всех трех уравнений.

Заметим, что объединение двух пар смежных уравнений, имеющих одно общее уравнение, может не быть смежным, так как третья пара из двух других уравнений не всегда является смежной.

Подсчитаем для $t = 2n$ число смежных пар уравнений.

Лемма 3. *Если* $t = 2n$, *то число разных смежных пар уравнений в системе* (6) *равно* n.

Доказательство. Для каждого $i \in \{1, 2, \ldots, k\}$ будет две пары смежных уравнений, содержащих уравнение S_i. Действительно, этими парами являются S_i, S_{i+k} и S_i, S_{i+n-k}. В первой паре общее неизвестное (ошибка) ε_{i+k}, а во

второй паре — ε_{i+n}. Если $i \in \{k+1, k+2, \ldots, n-k\}$, то существует одна смежная пара уравнений S_i, S_{i+k}. Для $i \in \{n-k+1, n-k+2, \ldots, 2n\}$ смежные пары не существуют. Следовательно, число смежных пар уравнений равно n.

Для $t = 2n$ смежная система из трех уравнений не существует. Поэтому будем рассматривать системы из трех уравнений, в которых две пары являются смежными.

Теорема 1. *Для смежной пары уравнений* (10) *с* $V_i + V_{i+k} = 1$ *справедливы следующие утверждения:*

1. *Если* $1 \le i \le k$, $V_i = 0$, $V_{i+k} = 1$, $V_{i+n-k} = 1$, *то* $\varepsilon_i = 0$.
2. *Если* $1 \le i \le n-2k$, $V_i = 1$, $V_{i+k} = 0$, $V_{i+2k} = 1$, *то* $\varepsilon_{i+n+k} = 0$.
3. *Если* $n-2k+1 \le i \le n-k$, $V_i = 1$, $V_{i+k} = 0$, $V_{i+2k-n} = 1$, *то* $\varepsilon_{i+n+k} = 0$.

Доказательство. Система из трех булевых уравнений

$$\varepsilon_i \oplus \varepsilon_{i+k} \oplus \varepsilon_{i+n} = 0,$$
$$\varepsilon_{i+k} \oplus \varepsilon_{i+2k} \oplus \varepsilon_{i+n+k} = 1,$$
$$\varepsilon_{i+n-k} \oplus \varepsilon_{i+n} \oplus \varepsilon_{i+2n-k} = 1$$

преобразуется в систему из четырех псевдобулевых уравнений

$$\varepsilon_{i+k} - \varepsilon_i - \varepsilon_{i+n} = 0,$$
$$\varepsilon_{i+n} - \varepsilon_i - \varepsilon_{i+k} = 0,$$
$$\varepsilon_{i+k} + \varepsilon_{i+2k} + \varepsilon_{i+n+k} = 1,$$
$$\varepsilon_{i+n-k} + \varepsilon_{i+n} + \varepsilon_{i+2n-k} = 1.$$

Действительно, первые два равенства получены по лемме 1 после рассмотрения двух смежных пар уравнений. Сложим первые два уравнения, получим $\varepsilon_i = 0$. Остальные два пункта доказываются аналогичным образом. По существу, мы здесь применили метод выделения и оценки отдельных неизвестных [4].

Следствие 3. *Согласно теореме* 1 о*шибки* ε_j *при* $j \in \{k+1, k+2, \ldots, n+k\}$ *не будут определены, а остальные* n *ошибок могут быть определены, но только по одному разу.*

Доказанная теорема является аналогом теоремы 2 из работы [3] для ошибок.

Лемма 4. *Среднее число* $N(t,n,k)$ *найденных по формулам теоремы* 1 *значений ошибок выражается формулой*

$$N(t,n,k) = \begin{cases} \frac{3}{64}(t-n+k), & \text{если } t > n+k, \\ \frac{3}{32}(t-n), & \text{если } t \le n+k. \end{cases}$$

Доказательство. Если $t-n>k$, то

$$N(t,n,k)=2k\cdot\frac{3}{64}+\frac{3}{64}\cdot(t-n-k)=\frac{3}{64}\cdot(t-n+k).$$

Для случая $t-n\le k$ получаем

$$N(t,n,k)=2(t-n)\cdot\frac{3}{64}=\frac{3}{32}\cdot(t-n).$$

Здесь мы воспользовались некоторыми результатами из работы [3].

Найдем такое t^*, для которого выполняются неравенства

$$N(t^*-1,n,k)<1,\ N(t^*,n,k)\ge 1.$$

Число уравнений $t=t^*$ можно считать необходимым для начала процесса частичного восстановления исходной рекурренты. Для $t>n+k$ число $N(t,n,k)\ge 1$ при $t\ge n-k+\frac{64}{3}$, а при $t\le n+k$ находим $t\ge n+\frac{32}{3}$. Объединяя эти два неравенства, получаем $t^*=n+11$ при $k\ge 11$ и $t^*=n+22-k$ для $k\le 10$.

Найдем вероятности событий $\{V_i=0\}$, $\{V_i=1\}$, $\{V_i=0, V_{i+k}=1\}$, полагая, что линейные формы $L_i, L_{i+k}, Q_i, Q_{i+k}$ линейно независимые и как случайные величины равновероятные. Очевидно, что

$$\mathrm{P}\{V_i=a_i\oplus a_{i+k}\oplus a_{i+n}=1\}=\mathrm{P}\{L_iQ_i\oplus L_{i+k}Q_{i+k}\oplus(L_i\oplus L_{i+k})(Q_i\oplus Q_{i+k})=1\}=$$

$$=3\,\mathrm{P}\{L_iQ_i=1, L_{i+k}Q_{i+k}\oplus(1\oplus L_{i+k})(1\oplus Q_{i+k})=1\}=\frac{6}{16}=\frac{3}{8}.$$

Аналогично находим

$$\mathrm{P}\{V_i=0,V_{i+k}=1\}=\frac{1}{4}\sum_{a,b=0}^{1}\mathrm{P}\{V_i=0,V_{i+k}=1, L_{i+k}=a,Q_{i+k}=b\}=$$

$$\frac{1}{4}\sum_{a,b=0}^{1}\mathrm{P}\{L_iQ_i\oplus(L_i\oplus a)(Q_i\oplus b)=ab,\ L_{i+2k}Q_{i+2k}\oplus(L_{i+2k}\oplus a)(Q_{i+2k}\oplus b)=1\oplus ab\}=$$

$$=\frac{1}{4}\sum_{a,b=0}^{1}\mathrm{P}\{bL_i\oplus aQ_i=0, bL_{i+2k}\oplus aQ_{i+2k}=1\}=\frac{3}{16}.$$

Далее, таким же образом получаем

$$\mathrm{P}\{V_i=1,V_{i+k}=1\}=\frac{3}{16},\ \mathrm{P}\{V_i=0,V_{i+k}=0\}=\frac{7}{16}.$$

$$\mathrm{P}\{V_{i+k}=0\,|\,V_i=1\}=\mathrm{P}\{V_{i+k}=0,V_i=1\}/\mathrm{P}\{V_i=1\}==\frac{3}{16}/\frac{3}{8}=\frac{1}{2},$$

$$\mathrm{P}\{V_{i+k}=1\,|\,V_i=0\}=\frac{3}{16}/\frac{5}{8}=\frac{3}{10},$$

$$\mathrm{P}\{V_{i+k}=1\,|\,V_i=1\}=\frac{3}{16}/\frac{3}{8}=\frac{1}{2},$$

$$\mathrm{P}\{V_{i+k}=0\,|\,V_i=0\}=\frac{7}{16}/\frac{5}{8}=\frac{7}{10}.$$

Приведем значения еще нескольких вероятностей:

$$\mathrm{P}\left(a_i=1,a_{i+k}=1,a_{i+n}=1\right)=6/2^6=\frac{1}{8}\cdot\left(1-\frac{1}{4}\right),$$

$$\mathrm{P}\left(a_i=1,a_{i+k}=1,a_{i+n}=0\right)=10/2^6=\frac{1}{8}\cdot\left(1+\frac{1}{4}\right),$$

$$\mathrm{P}\left(a_i=0,a_{i+k}=0,a_{i+n}=1\right)=6/2^6=\frac{1}{8}\cdot\left(1-\frac{1}{4}\right),$$

$$\mathrm{P}\left(a_i=0,a_{i+k}=0,a_{i+n}=0\right)=10/2^6=\frac{1}{8}\cdot\left(1+\frac{1}{4}\right).$$

Значения a_i,a_{i+k},a_{i+n} в круглых скобках можно переставлять. Отсюда, в частности, следует, что

$$\mathrm{P}\left(a_{i+n}=0\,|\,a_i=0,a_{i+k}=0\right)=\frac{1}{2}\cdot\left(1+\frac{1}{4}\right),$$

$$\mathrm{P}\left(a_{i+n}=0\,|\,a_i=0,a_{i+k}=1\right)=\frac{1}{2}\cdot\left(1-\frac{1}{4}\right).$$

Эти результаты можно учитывать при построении критериев различения случайной системы уравнений и заведомо совместной случайной системы уравнений. Если мы не знаем значения k,n, то для истинной пары (k,n) выполняются вышеприведенные равенства.

Теорема 2. *Для каждой смежной пары уравнений*

$$\varepsilon_i\oplus\varepsilon_{i+k}\oplus\varepsilon_{i+n}=V_i,$$

$$\varepsilon_{i+k}\oplus\varepsilon_{i+2k}\oplus\varepsilon_{i+n+k}=V_{i+k},$$

правые части которой удовлетворяют условию $V_i + V_{i+k} = 1$, *при* $i \le t - 2n$ *справедливы следующие утверждения:*

1. *Если* $V_{i+n} = 0$, *то* $\varepsilon_{i+n} = 0$ *при* $V_i = 0$ *и* $\varepsilon_{i+n+k} = 0$ *при* $V_i = 1$.
2. *Если* $V_{i+n} = 1$, *то* $\varepsilon_{i+n+k} = 0$ *при* $V_i = 1$ *и* $\varepsilon_i = 0$ *при* $V_i = 0$.
3. *Если* $W_i = 0$, *то* $\varepsilon_i = 0$ *при* $V_i = 0$ *и* $\varepsilon_{i+2k} = 0$ *при* $V_i = 1$.

Если $W_i = 1$, *то* $\varepsilon_{i+n} = 0$ *при* $V_i = 0$ *и* $\varepsilon_{i+n+k} = 0$ *при* $V_i = 1$.

Доказательство. Аналогично доказательству теоремы 1 систему из трех смежных уравнений

$$\varepsilon_i \oplus \varepsilon_{i+k} \oplus \varepsilon_{i+n} = V_i = 0,$$

$$\varepsilon_{i+k} \oplus \varepsilon_{i+2k} \oplus \varepsilon_{i+n+k} = V_{i+k} = 1,$$

$$\varepsilon_i \oplus \varepsilon_{i+2k} \oplus \varepsilon_{i+2n} = W_i = 0$$

заменяем системой из псевдобулевых уравнений

$$\varepsilon_{i+k} - \varepsilon_i - \varepsilon_{i+n} = 0,$$

$$\varepsilon_{i+k} + \varepsilon_{i+2k} + \varepsilon_{i+n+k} = 1,$$

$$\varepsilon_{i+2k} - \varepsilon_i - \varepsilon_{i+2n} = 0.$$

Сложим первое и третье уравнения:

$$\varepsilon_{i+k} + \varepsilon_{i+2k} - 2\varepsilon_i - \varepsilon_{i+n} - \varepsilon_{i+2n} = 0.$$

Так как $\varepsilon_{i+k} + \varepsilon_{i+2k} \le 1$, то $\varepsilon_i = 0$, $\varepsilon_{i+n} + \varepsilon_{i+2n} \le 1$. Для $V_i = 1, V_{i+k} = 0$ находим $\varepsilon_{i+2k} = 0$.

Аналогично доказываются остальные пункты теоремы и следующее утверждение.

Теорема 3. *Для каждой смежной пары уравнений* (10) *при* $V_i = V_{i+k}, i \le t - 2n$ *справедливы следующие утверждения:*

1. *Если* $V_i = V_{i+k} = 1, V_{i+n} = 0$, *то* $\varepsilon_{i+2n} = 0$.
2. *Если* $V_i = V_{i+k} = 0, V_{i+n} = 1$, *то* $\varepsilon_{i+k} = 0$.
3. *Если* $W_i = 0, V_i = V_{i+k} = 1$, *то* $\varepsilon_{i+2n} = 0$.
4. *Если* $W_i = 1, V_i = V_{i+k} = 0$, *то* $\varepsilon_{i+k} = 0$.

Перечислим случаи, когда мы не сможем определить значение одной ошибки. Для этого должны осуществиться следующие события:

$$V_i = V_{i+k} = V_{i+n} = W_i \text{ (теорема 3).}$$

Для случайной системы уравнений эта вероятность равна $1/4$. Далее, необходимо, чтобы не осуществились события, описанные в теореме 1 для $t \le 2n$.

По существу эти две теоремы утверждают, что ошибки всегда определяются при выполнении неравенств

$$1 \le V_i + V_{i+k} + V_{i+n} \le 2 .$$

Отсюда следует вывод, что, во-первых, эти два утверждения можно объединить, а, во-вторых, следует отдельно рассмотреть два случая

$$V_i = V_{i+k} = V_{i+n} = a, a \in \{0,1\} .$$

Первый случай ($a = 0$) сводится к решению системы из трех булевых уравнений относительно 6 неизвестных, число решений равно 8. Второй случай ($a = 1$) приводит нас к необходимости решать систему из трех псевдобулевых уравнений

$$\varepsilon_i + \varepsilon_{i+k} + \varepsilon_{i+n} = 1 ,$$

$$\varepsilon_{i+k} + \varepsilon_{i+2k} + \varepsilon_{i+n+k} = 1 ,$$

$$\varepsilon_{i+n} + \varepsilon_{i+n+k} + \varepsilon_{i+2n} = 1 .$$

Эта система имеет 3 решения относительно 6 неизвестных. Число ошибок, равных 1, в трех решениях равно 2.

Приведем более общее утверждение, не связанное с ограничением на значение t . Пусть

$$R = \left\{ \left(i, i + 2^s k, i + 2^s n \right), s \ge 0, i = 1,2,\ldots \right\} .$$

Теорема 4. *Если каждая из трех троек* $\left(j_1, j_2, j_3 \right), \left(j_1, j_4, j_5 \right), \left(j_4, j_6, j_7 \right)$ *после некоторой перестановки координат попадает в множество* R *, то из системы уравнений*

$$\varepsilon_{j_1} \oplus \varepsilon_{j_2} \oplus \varepsilon_{j_3} = a\left(j_1, j_2, j_3 \right) ,$$

$$\varepsilon_{j_1} \oplus \varepsilon_{j_4} \oplus \varepsilon_{j_5} = a\left(j_1, j_4, j_5 \right) ,$$

$$\varepsilon_{j_4} \oplus \varepsilon_{j_6} \oplus \varepsilon_{j_7} = a\left(j_4, j_6, j_7 \right)$$

следует, что $\varepsilon_{j_5} = 0$ *при* $a\left(j_1, j_2, j_3 \right) = a\left(j_4, j_6, j_7 \right) = 1, a\left(j_1, j_4, j_5 \right) = 0$. *Если*

$$\left\{ j_2, j_3 \right\} \cap \left\{ j_6, j_7 \right\} = \left\{ j_2 \right\} \left(\left\{ j_3 \right\} \right) ,$$

$$a\left(j_1, j_2, j_3 \right) = a\left(j_4, j_6, j_7 \right) = 0, a\left(j_1, j_4, j_5 \right) = 1 ,$$

то $\varepsilon_{j_2} = 0 \left(\varepsilon_{j_3} = 0 \right)$.

4. Восстановление линейной рекуррентной последовательности по известным вариантам значений для каждого члена рекурренты

Пусть последовательность (1) удовлетворяет рекуррентному соотношению

$$x_i^0 + x_{i+k}^0 + x_{i+n}^0 = 0$$

над полем $GF(q), q \geq 3$. Числа k и n нам неизвестны, но известно, что выполняется система включений $x_j \in A_j$, $|A_j| = m < q$, $j = 1,2,\ldots,t$. Для заведомо совместной случайной системы включений истинное значение $x_j^0 \in A_j$, остальные $m-1$ знаков из множества A_j являются выборкой без возвращения из оставшихся $q-1$ знаков, все выборки независимы и равновероятны, $j = 1,2,\ldots,t$. В этом случае для истинных значений k и n значения членов рекурренты принадлежат соответственно множествам $A_1, A_2, \ldots, A_t$.

Для случайной системы включений считаем, что m знаков из множества A_j, $j = 1,2,\ldots,t$, являются равновероятной выборкой без возвращения из q элементов поля $GF(q)$. Все выборки независимы; поэтому для любого вектора $\overline{x} = (x_1, x_2, \ldots, x_n)$ и любых значений $1 \leq k < n \leq N$

$$\mathrm{P}\left(x_j \in A_j,\ j = n+1, n+2, \ldots, t\right) = \left(\frac{m}{q}\right)^{t-n},$$

а среднее число вариантов $k, n, \overline{x}$, удовлетворяющих условиям $x_j \in A_j, j = 1,2,\ldots,t$, не превосходит величины

$$\left(\frac{m}{q}\right)^t \sum_{3 \leq n \leq N} nq^n .$$

Очевидно, что

$$\sum_{3 \leq n \leq N} nq^n = q\left(\sum_{3 \leq n \leq N} x^n\right)'_{x=q} < q\frac{Nq^{N+1} - (N+1)q^N}{(q-1)^2} < Nq^N \frac{q}{q-1}.$$

Отсюда находим, что среднее число вариантов значений $k, n, \overline{x}$, удовлетворяющих указанным условиям, стремится к нулю, если

$$\left(\frac{m}{q}\right)^t \cdot Nq^N \to 0$$

при $t, N \to \infty$ и фиксированном q.

Следовательно, мы приходим к следующему утверждению.

Теорема 5. *Если* $m < q$, q *фиксировано,* $t, N \to \infty$ *и*

$$\liminf_{t,N\to\infty} \frac{t(\ln q - \ln m)}{N \ln q} > 1,$$

то случайная система включений $X_j \in A_j$, $j = 1,2,\ldots,t$, *не имеет решений с вероятностью, стремящейся к единице.*

Заметим, что в некоторых случаях можно отказаться от перебора всех q^n возможных начальных заполнений исходной рекурренты и ограничиться перебором значений k, n и небольшой части начального заполнения. Например, можно задать $x_i, x_{i+k}, \ldots, x_{i+sk}$, найти $x_{i+n}, x_{i+n+k}, \ldots, x_{i+(s-1)k}, x_{i+2n}, \ldots$, проверить соотношения $x_{i+n} \in A_{i+n}, x_{i+n+k} \in A_{i+n+k}, \ldots$ Если соотношения не выполняются, то переходим к другой паре значений k, n. В ином случае задаем $x_{i+(s+1)k}$ и проверяем, выполняются ли новые соотношения, и т. д.

5. Частичное восстановление последовательности по наблюдениям над k-граммами

Пусть исходная последовательность (1) состоит из знаков, принимающих значения $0,1,\ldots,q-1$. В этой последовательности могут встречаться только k-граммы из некоторого известного множества B, $|B| = n$, $1 \le n < q^k$. Множество B может быть множеством решений некоторого уравнения

$$F(x_i, x_{i+1}, \ldots, x_{i+k-1}) = 0.$$

Имеются наблюдения над знаками, знак x_i может принимать только значения из множества знаков A_i, $|A_i| = s$, $1 \le s < q$. Для истинной последовательности $x_1^0, x_2^0, \ldots, x_t^0$ выполняются условия

$$x_i^0 \in A_i, \; i = 1,2,\ldots,t,$$

$$\left(x_i^0, x_{i+1}^0, \ldots, x_{i+k-1}^0\right) \in B, i = 1,2,\ldots,t-k+1.$$

Необходимо найти такие условия на t,k,q,s и множество B, при которых можно определить несколько участков исходной последовательности.

Пусть $n_i(j)$ есть число k-грамм из множества B, отличающихся i знаками от истинной k-граммы, находящейся на j-м месте. Очевидно, что

$$\sum_{i=0}^{k} n_i(j) = n, \quad n_0(j) = 1, \quad j = 1,2,\ldots,t-k+1.$$

Набор этих чисел назовем характеристикой множества B на j-м месте k-граммы. Обозначим $B(j)$ множество допустимых k-грамм на j-м месте после учета наблюдений над знаками этой k-граммы, $B(j) \subset B$.

Оценим $|B(j)|$ и характеристику множества $B(j)$, полагая, что во множество A_i отбираются $s-1$ знаков случайно и равновероятно без возвращения из совокупности $A_0 \setminus \{x_i^0\}$, $A_0 = \{0,1,\ldots,q-1\}$. Характеристика множества $B(j)$ состоит из набора случайных величин

$$\{\xi_0(j), \xi_1(j), \ldots, \xi_k(j)\},$$

где $\xi_i(j)$ — число k-грамм из множества $B(j)$, отличных от истинной k-граммы на j-м месте в i местах, $\xi_0(j) \equiv 1$. Нетрудно увидеть, что

$$\mathrm{M}\xi_i(j) = n_i(j) \cdot \left(\frac{s-1}{q-1}\right)^i. \tag{11}$$

Следовательно,

$$\mathrm{M}|B(j)| = 1 + \sum_{i=1}^{k} n_i(j) \cdot \left(\frac{s-1}{q-1}\right)^i =$$

$$= 1 - \left(\frac{s-1}{q-1}\right)^k + \sum_{i=1}^{k} n_i(j) \cdot \left[\left(\frac{s-1}{q-1}\right)^i - \left(\frac{s-1}{q-1}\right)^k\right] + n \cdot \left(\frac{s-1}{q-1}\right)^k.$$

Из приведенных выше формул можно оценить значения t,k,q,s.

Среди $t-k+1$ величин $|B(j)|$ найдем наименьшую величину $|B(l)|$. Рассмотрим далее соседние k-граммы на $(l-1)$-м месте и $(l+1)$-м месте

и найдем множество возможных $(k+2)$-грамм $x_{l-1}, x_l, \ldots, x_{l+k-1}, x_{l+k}$. В этом множестве выделим подмножество возможных k-грамм $x_l, x_{l+1}, \ldots, x_{l+k-1}$, оно будет входить в множество $B(l)$. Этот процесс продолжим до однозначного нахождения истинной k-граммы $x_l^0, x_{l+1}^0, \ldots, x_{l+k-1}^0$.

Второй способ связан с учетом известной характеристики множества B на j-м участке, так как характеристика множества $B(j)$ связана статистически с этой характеристикой (см. формулу (11)). Для этого мы перебираем все варианты из $B(j)$ и находим для каждого варианта его предполагаемую характеристику. То решение, характеристика которого лучше согласуется с формулой (11), можно объявить претендентом на истинное решение и попробовать его продолжить справа и слева.

Список литературы

1. *Чистяков, В. П.* Курс теории вероятностей. 5-е издание. — М.: Агар, 2000. — 256 с.
2. *Балакин, Г. В.* Системы булевых уравнений с искаженной правой частью при ограничениях на значения неизвестных и ошибок // В сб.: Труды по дискретной математике. — Т. 11. Вып. 1. — М.: ФИЗМАТЛИТ, 2008. — С. 5–17.
3. *Балакин, Г. В.* О решении некоторых классов систем булевых уравнений рекуррентного типа // Математические вопросы криптографии. — 2013. — Т. 4. Вып. 1. — С. 17–38.
4. *Балакин, Г. В.* О возможности решения систем линейных целочисленных уравнений методом выделения и оценки отдельных неизвестных // Дискретная математика. — 1994. — Т. 6. Вып. 1. — С. 116–126.