ПРИКЛАДНАЯ ТЕОРИЯ КОДИРОВАНИЯ

УДК 621.391.7

ПРИМЕНЕНИЕ ОДНОГО МЕТОДА РАСПОЗНАВАНИЯ ЛИНЕЙНОГО КОДА ДЛЯ КАНАЛА С ПОДСЛУШИВАНИЕМ

Ю. В. Косолапов, О. Ю. Турченко

Южный федеральный университет, г. Ростов-на-Дону, Россия

Рассматривается модель защиты данных с помощью метода кодового зашумления. Предполагается, что кодируемые информационные блоки длины k содержат фиксированное сообщение $\mathbf{m}$ длины m_l ($m_l \leqslant k$) на фиксированной позиции m_s ($1 \leqslant m_s \leqslant k - m_l + 1$), а наблюдатель получает зашумлённые кодовые слова длины n через q-ичный (q — степень простого числа) симметричный канал с вероятностью p для каждого ненулевого значения ошибки. Целью наблюдателя является нахождение сообщения $\mathbf{m}$, когда позиция m_s и длина m_l неизвестны. Предложен способ нахождения сообщения $\mathbf{m}$ и получена оценка количества наблюдаемых кодовых слов, достаточного для восстановления сообщения $\mathbf{m}$ этим способом.

Ключевые слова: *кодовое зашумление, q-ичный симметричный канал, распознавание кода.*

DOI 10.17223/20710410/35/7

APPLICATION OF ONE METHOD OF LINEAR CODE RECOGNITION TO THE WIRE-TAP CHANNEL

Y. V. Kosolapov, O. Y. Turchenko

Southern Federal University, Rostov-on-Don, Russia

E-mail: itaim@mail.ru

A security model using the method of code noising is considered. It is assumed that the information blocks of length k contain a fixed message $\mathbf{m}$ of length m_l ($m_l \leqslant k$) on a fixed position m_s ($1 \leqslant m_s \leqslant k - m_l + 1$), and an observer gets noisy codewords of length n through a q-ary symmetric channel $q\mathrm{SC}(p)$ (q — prime power) with error probability p ($p < 1/q$) for each non-zero value. The aim of the observer is to find the unknown message $\mathbf{m}$, when the position m_s and the length m_l are unknown. In this paper, we propose a statistical method for finding message $\mathbf{m}$. The method is based on the idea of code recognition in noisy environment: we test hypothesis H_0 (received vectors have been generated by a conjectural code $\mathcal{C}$) against the hypothesis H_1 (received vectors have not been generated by $\mathcal{C}$ or it's subcode). The method is as follows. From the observed vectors, the independent pairwise differences of them are compiled. The resulting vectors are code words of some unknown code $\mathcal{C}$ noised in a symmetric channel $q\mathrm{SC}(p(2 - qp))$. To determine the length m_l and place m_s of the unknown message $\mathbf{m}$, we recognize the code $\mathcal{C}$ from the calculated differences of the received vectors. For this purpose, we present a code recognition algorithm (called CodeRecognition) and prove that if $\mathcal{C}$ is a linear $[n, k]$ code and $M(\mathcal{C}, \alpha, \beta)$ is

the minimum number of vectors (received from the channel $q\mathrm{SC}(p)$) that are sufficient for CodeRecognition algorithm to test the hypothesis H_0 against the hypothesis H_1 by using a constructed statistical criterion, then $M(\mathcal{C}, \alpha, \beta) \leqslant f^2(k+1)$, where

$$f(x) = \frac{b(1+(q-2)(1-pq)^x - (q-1)(1-pq)^{2x})^{1/2} - a}{\sqrt{q-1}(1-pq)^x},$$

α and β are the probabilities of errors of the first kind and the second kind, respectively; $a = \Phi^{-1}(\alpha)$, $b = \Phi^{-1}(1-\beta)$; Φ^{-1} — Laplacian inverse function. We show that the bound above is achieved in the case of Maximum Distance Separable (MDS) codes $\mathcal{C}$. On the base of this result, we obtain a sufficient number of vectors corresponding to the channel $q\mathrm{SC}(p(2-qp))$. We also present some algorithms for finding the position m_s, the length m_l, and the message $\mathbf{m}$. The main component of them is the algorithm CodeRecognition.

Keywords: *code noising, q-ary symmetric channel, code recognition.*

Введение

Рассмотрим информационно-аналитическую систему (ИАС), в которой два легальных участника (отправитель и получатель) связаны каналом без помех, а пассивный наблюдатель подслушивает передаваемые данные по q-ичному (q — степень простого числа) симметричному каналу с вероятностью p для каждого ненулевого значения ошибки. Эта модель является частным случаем модели канала с наблюдением, впервые рассмотренной в [1]. Как и в [1], предполагается, что отправитель для защиты от наблюдения использует метод кодирования смежными классами (метод кодового зашумления). У наблюдателя при подслушивании одного кодового слова из-за наличия помех возникает неопределённость относительно сообщения, которое было закодировано. Вычислению этой неопределённости в неасимптотическом случае для двоичного симметричного канала наблюдения (при $q = 2$) посвящена работа [2], а в [3] показано, что эта неопределённость может быть снята в рамках модели многократной передачи данных. В частности, в [3] найдена оценка количества зашумлённых кодовых слов, соответствующих одному информационному блоку, достаточного для нахождения этого блока с заданными вероятностями ошибок первого и второго рода. Отметим, что на метод определения информационного блока, основанный на результатах работы [3], накладывается ограничение: наблюдаемые векторы должны соответствовать одному и тому же информационному блоку. В настоящей работе рассматривается более общая задача нахождения информационного блока в рамках модели многократной передачи данных. Предполагается, что различные кодируемые информационные блоки содержат фиксированное сообщение на фиксированной позиции, а наблюдатель получает зашумлённые кодовые слова через q-ичный симметричный канал. В работе ставится задача нахождения позиции и длины подвектора. Зная эти параметры, содержание подвектора может быть определено, например, с помощью метода, основанного на результатах работы [3]. Отметим, что для случая $q = 2$ задача в такой постановке частично решена [4]. Рассмотрение q-ичного симметричного канала представляет не только теоретический интерес. Эта модель канала, например, в последнее время исследуется как адекватная модель пакетной передачи данных [5], а также как модель генерации ошибок в протоколах гомоморфного шифрования [6].

Работа имеет следующую структуру: п. 1 посвящён детальному описанию информационно-аналитической системы. В п. 2 описан метод, разработанный для решения поставленной задачи. Этот метод описан в общем виде, так как может быть применён

для решения других задач, например задач типа распознавания кода. В п. 3 получены оценки применимости разработанного метода в рамках поставленной задачи и построены соответствующие алгоритмы.

1. Информационно-аналитическая система

Пусть информационными блоками являются векторы длины k над полем Галуа $\mathbb{F}_q$ мощности q, где q — степень простого числа. Предполагается, что в момент времени $t \in \mathbb{N}$ информационный блок $\mathbf{s}^{(t)} \in \mathbb{F}_q^k$ имеет вид $\mathbf{s}^{(t)} = (\mathbf{m}_1^{(t)} \parallel \mathbf{m} \parallel \mathbf{m}_2^{(t)})$, где подвектор $\mathbf{m} \in \mathbb{F}_q^{m_l}$ постоянный для всех t, а подвекторы $\mathbf{m}_1^{(t)}$ и $\mathbf{m}_2^{(t)}$ распределены случайно и равновероятно со значениями из $\mathbb{F}_q^{m_s-1}$ и $\mathbb{F}_q^{k-[m_l+m_s]+1}$ соответственно; запись $(\mathbf{a} \parallel \mathbf{b})$ здесь и далее обозначает конкатенацию векторов $\mathbf{a}$ и $\mathbf{b}$. Перед передачей в канал информационные сообщения кодируются отправителем с помощью метода кодового зашумления [2], а именно: для зафиксированных натуральных чисел k и n $(k < n)$ легальными участниками выбирается $((n-k)\times k)$-матрица P с элементами из поля $\mathbb{F}_q$, а каждое сообщение $\mathbf{s}^{(t)}$ длины k кодируется по следующему правилу:

$$\mathrm{Enc}(\mathbf{s}^{(t)}) = (\mathbf{s}^{(t)} + \mathbf{K}^{(t)}P \parallel \mathbf{K}^{(t)}) = \mathbf{c}^{(t)} = (\mathbf{c}_1^{(t)} \parallel \mathbf{K}^{(t)}), \tag{1}$$

где $\mathbf{K}^{(t)}$ — случайно и равновероятно выбранный вектор из $\mathbb{F}_q^{n-k}$. Матрица P и правило кодирования (1) известны всем участникам ИАС (в том числе и наблюдателю), поэтому при отсутствии помех в канале между отправителем и получателем правило декодирования имеет вид $\mathrm{Dec}(\mathbf{c}^{(t)}) = \mathbf{K}^{(t)}P + \mathbf{c}_1^{(t)} = \mathbf{s}^{(t)}$.

Предположим, что наблюдатель передаваемые кодовые слова подслушивает через q-ичный симметричный канал, в котором вероятность возникновения ошибки в каждом символе равна $(q-1)p$, $0 < p < 1/q$, а именно: если $\mathbf{c}^{(t)}$ — вектор длины n на входе канала наблюдения, а $\mathbf{z}^{(t)} = \mathbf{c}^{(t)} + \mathbf{e}^{(t)}$ — вектор на выходе этого канала, то значения координат вектора ошибок $\mathbf{e}^{(t)} = (e_1^{(t)}, \ldots, e_n^{(t)})$ принимают значения в соответствии со следующим распределением вероятностей:

$$\Pr\{\mathbf{e}_i^{(t)} = l\} = \begin{cases} p, & l \in \mathbb{F}_q^* = \mathbb{F}_q \setminus \{0\}, \\ 1-(q-1)p, & l = 0, \end{cases} \qquad i \in \{1, \ldots, n\}.$$

В дальнейшем такой q-ичный симметричный канал будем обозначать $q\mathrm{SC}(p)$. Наблюдателю позиции m_s подвектора $\mathbf{m}$ и его длина m_l неизвестны. Целью наблюдателя является нахождение неизвестного подвектора $\mathbf{m}$ при многократной передаче сообщений $\mathbf{s}^{(t)}$, закодированных по правилу (1). Естественно полагать, что эта задача решается наблюдателем последовательно: сначала находятся неизвестные позиция m_s и длина m_l, а затем само сообщение $\mathbf{m}$. Отметим, что при известных m_s и m_l задача нахождения содержания вектора $\mathbf{m}$ может быть решена методом, основанным на результатах работы [3]. Поэтому ниже строится способ нахождения неизвестных значений m_s и m_l.

Наблюдателем выдвигается гипотеза $H_{i,j}$ о том, что $m_s = i$ и $m_l \geqslant j$ (т. е. предполагаемая длина j не превышает истинного значения длины сообщения). Далее для обозначения векторов и матриц в рамках данной гипотезы будем использовать верхний индекс (i,j). В этом случае матрица P представима в блочном виде: $P = [P_1^{(i,j)}|P_2^{(i,j)}|P_3^{(i,j)}]$, где $P_1^{(i,j)}$ — первые $i-1$ столбцов матрицы P; $P_2^{(i,j)}$ — столбцы матрицы P с номерами от i до $i+j-1$; $P_3^{(i,j)}$ — последние $k-(i+j)+1$ столбцов матрицы P. В рамках этой гипотезы наблюдаемый в момент времени t вектор $\mathbf{z}^{(t)} = \mathbf{c}^{(t)} + \mathbf{e}^{(t)}$

имеет следующий вид:

$$\mathbf{z}^{(t)} = (\widehat{\mathbf{m}}_1^{(t)} + \mathbf{K}^{(t)} P_1^{(i,j)} + \mathbf{e}_1^{(t)} \parallel \widehat{\mathbf{m}} + \mathbf{K}^{(t)} P_2^{(i,j)} + \mathbf{e}_2^{(t)} \parallel \widehat{\mathbf{m}}_2^{(t)} + \mathbf{K}^{(t)} P_3^{(i,j)} + \mathbf{e}_3^{(t)} \parallel \mathbf{K}^{(t)} + \mathbf{e}_4^{(t)}), \quad (2)$$

где $\mathbf{e}^{(t)} = (\mathbf{e}_1^{(t)} \parallel \mathbf{e}_2^{(t)} \parallel \mathbf{e}_3^{(t)} \parallel \mathbf{e}_4^{(t)})$ — вектор помехи в канале; $\mathbf{e}_1^{(t)} \in \mathbb{F}_q^{i-1}$; $\mathbf{e}_2^{(t)} \in \mathbb{F}_q^{j}$; $\mathbf{e}_3^{(t)} \in \mathbb{F}_q^{k-(i+j)+1}$; $\mathbf{e}_4^{(t)} \in \mathbb{F}_q^{n-k}$; символ «$\widehat{\ }$» означает, что соответствующие подвекторы могут отличаться от истинных, если гипотеза неверна.

Для подмножества $S = \{i_1, \ldots, i_u\}$ множества $\{1, \ldots, n\}$ рассмотрим линейный оператор проекции $\pi_S : \mathbb{F}^n \to \mathbb{F}^u$, ставящий в соответствие каждому вектору $\mathbf{x} = (x_1, \ldots, x_n)$ из $\mathbb{F}^n$ вектор $\mathbf{x}_S = (x_{i_1}, \ldots, x_{i_u})$. Рассмотрим множество координат $\tau(i,j) = \{i, \ldots, i+j-1\} \cup \{k+1, \ldots, n\}$ и выборку из N векторов (N чётное)

$$\mathbf{Z}_{\tau(i,j)} = \left(\mathbf{z}_{\tau(i,j)}^{(1)}, \mathbf{z}_{\tau(i,j)}^{(2)}, \ldots, \mathbf{z}_{\tau(i,j)}^{(N)}\right), \quad (3)$$

где $\mathbf{z}_{\tau(i,j)}^{(t)} = \pi_{\tau(i,j)}(\widehat{\mathbf{z}}^{(t)})$ — проекция наблюдаемого вектора $\mathbf{z}^{(t)}$ на множество координат $\tau(i,j)$. По выборке (3) построим набор векторов мощности $N/2$:

$$\widetilde{\mathbf{Z}} = (\widetilde{\mathbf{z}}^{(1)}, \widetilde{\mathbf{z}}^{(2)}, \ldots, \widetilde{\mathbf{z}}^{(N/2)}). \quad (4)$$

Здесь $\widetilde{\mathbf{z}}^{(t)} = \mathbf{z}_{\tau(i,j)}^{(2t)} - \mathbf{z}_{\tau(i,j)}^{(2t-1)}$, $t = 1, \ldots, N/2$. С учётом (2), вектор $\widetilde{\mathbf{z}}^{(t)}$ в выборке (4) представим в виде

$$\begin{aligned} \widetilde{\mathbf{z}}^{(t)} &= \left((\mathbf{K}^{(2t)} - \mathbf{K}^{(2t-1)}) P_2^{(i,j)} + \mathbf{e}_2^{(2t)} - \mathbf{e}_2^{(2t-1)} \parallel \mathbf{K}^{(2t)} - \mathbf{K}^{(2t-1)} + \mathbf{e}_4^{(2t)} - \mathbf{e}_4^{(2t-1)}\right) = \\ &= \left(\mathbf{K}^{(2t)} - \mathbf{K}^{(2t-1)}\right) [P_2^{(i,j)} \parallel I_{n-k}] + \left(\mathbf{e}_2^{(2t)} - \mathbf{e}_2^{(2t-1)} \parallel \mathbf{e}_4^{(2t)} - \mathbf{e}_4^{(2t-1)}\right), \end{aligned} \quad (5)$$

где I_{n-k} — единичная матрица. Далее линейный код $\mathcal{C}$ размерности k и длины n будем называть, как обычно, $[n,k]$-кодом [7], а его порождающую матрицу обозначим $G(\mathcal{C})$. Фиксированный набор базисных векторов кода $\mathcal{C}$ обозначим $B_{\mathcal{C}}$, а множество всех различных базисов кода $\mathcal{C}$ — символом $\mathcal{B}_{\mathcal{C}}$.

Лемма 1. Пусть выборка $\mathbf{Z}_{\tau(i,j)}$ вида (3) построена по векторам вида (2), наблюдаемым на выходе $q\mathrm{SC}(p)$; $\mathcal{C}^{(i,j)}$ — $[n-k+j, n-k]$-код с порождающей матрицей $G(\mathcal{C}^{(i,j)}) = [P_2^{(i,j)} | I_{n-k}]$; $H_{i,j}$ — выдвигаемая гипотеза.

1) Если $i = m_s$ и $j \leqslant m_l$, то набор $\widetilde{\mathbf{Z}}$ вида (4) представляет собой набор из зашумлённых кодовых слов кода $\mathcal{C}^{(i,j)}$, полученных из $q\mathrm{SC}(\widehat{p})$, где $\widehat{p} = p(2-pq)$.
2) Пусть множество координат $\{i, \ldots, i+j-1\}$ не вложено в $\{m_s, \ldots, m_s + m_l - 1\}$. Тогда вероятность того, что выборка $\widetilde{\mathbf{Z}}$ представляет собой набор из зашумлённых кодовых слов кода $\mathcal{C}^{(i,j)}$, полученных из $q\mathrm{SC}(\widehat{p})$, не превосходит $(1/q)^{N/2}$.

Доказательство. Так как, по предположению, гипотеза верная, то из (5) следует, что набор (4) является набором из зашумлённых кодовых слов линейного кода $\mathcal{C}^{(i,j)}$ с порождающей матрицей $G(\mathcal{C}^{(i,j)}) = [P_2^{(i,j)} | I_{n-k}]$.

Покажем, что $\Pr\{e_i = l\} = \widehat{p}$ для всех $i \in \{1, \ldots, n\}$, $l \in \mathbb{F}_q^*$. Поскольку в (5) участвует разность помех, реализованных в канале $q\mathrm{SC}(p)$, то по формуле полной вероятности вероятность $\Pr\{e_i = l\}$ для ненулевого l равна сумме вероятностей всевозможных разностей помех, которые в результате дают l. Таких разностей ровно q, причём разностей, содержащих нулевое значение, ровно две: $l-0$ и $0-(q-l)$. Остальные разности

содержат ненулевые значения в качестве уменьшаемого и вычитаемого, которые равновероятны (с вероятностью p), поэтому каждая пара реализуется с вероятностью p^2, их количество равно $q-2$. Таким образом,

$$\widehat{p} = 2(1-(q-1)p)p + \sum_{m=1}^{q-2} p^2 = 2p - 2(q-1)p^2 + p^2(q-2) = p(2-pq).$$

Докажем второе утверждение леммы. Пусть R — множество координат из $\{i, \ldots, i+j-1\}$, которые отсутствуют в $\{m_s, \ldots, m_s + m_l - 1\}$. Тогда вектор из выборки (4) представляет собой сумму зашумлённого кодового слова кода $\mathcal{C}^{(i,j)}$ и случайного равновероятного вектора $\mathbf{g}$ веса $|R|$ (координаты этого вектора представляют собой координаты разности векторов $\mathbf{m}_1^{(2t)} - \mathbf{m}_1^{(2t-1)}$ или $\mathbf{m}_2^{(2t)} - \mathbf{m}_2^{(2t-1)}$). Таким образом, для того чтобы вектор из выборки (4) представлял собой зашумлённое кодовое слово кода $\mathcal{C}^{(i,j)}$, вектор $\mathbf{g}$ должен быть нулевым. Вероятность этого события для одного вектора равна $(1/q)^{|R|}$. Тогда для всей выборки вероятность соответствующего события есть $(1/q)^{|R|N/2}$. Наибольшее значение этой вероятности достигается при $|R| = 1$. ■

Отметим, что функция $g(x) = x(2-xq)$ при каждом натуральном $q \geqslant 2$ возрастает от 0 до $1/q$ на отрезке $[0, 1/q]$. Следовательно, набор (4) представляет собой набор кодовых слов, более зашумлённых, чем слова из набора (3).

В силу леммы 1, проверка верности гипотез наблюдателем сводится к задаче распознавания кода по зашумлённому набору векторов. Заметим, что эта задача имеет несколько способов решения [8, 9]. В настоящей работе применяется метод, схожий с методом из работы [8], идея которого состоит в том, что вес синдрома зашумлённого в двоичном симметричном канале ($q = 2$) кодового слова в среднем меньше веса синдрома произвольного (случайно выбранного) вектора. Обобщённый метод для любого допустимого q приведён далее.

2. Метод распознавания кода в канале $q\mathrm{SC}(p)$

Рассмотрим линейный $[n, k, d]$-код $\mathcal{C}$ ($[n, k]$-код с кодовым расстоянием d) и ему дуальный $[n, n-k]$-код $\mathcal{C}^\perp$. Запись $(\mathbf{a}, \mathbf{b})$ обозначает скалярное произведение векторов $\mathbf{a}$ и $\mathbf{b}$, а для произвольного вектора $\mathbf{h} \in \mathbb{F}_q^n$ символ $\mathbf{h}^\perp$ обозначает подпространство размерности $n-1$, ортогональное вектору $\mathbf{h}$.

Лемма 2. Пусть $\mathbf{h} \in \mathbb{F}_q^n$, $\mathbf{c} \in \mathbf{h}^\perp$, $\mathbf{z} = \mathbf{c} + \mathbf{e}$ — вектор на выходе $q\mathrm{SC}(p)$. Тогда

$$\Pr\{(\mathbf{z}, \mathbf{h}) = 0\} = q^{-1}(1 + (q-1)(1-pq)^{\mathrm{w}(\mathbf{h})}), \tag{6}$$

где $\mathrm{w}(\mathbf{h})$ — вес Хэмминга вектора $\mathbf{h}$.

Доказательство. Справедлива следующая цепочка равенств:

$$\begin{gathered}\Pr\{(\mathbf{z}, \mathbf{h}) = 0\} = \Pr\{(\mathbf{c}+\mathbf{e}, \mathbf{h}) = 0\} = \Pr\{(\mathbf{e}, \mathbf{h}) = 0\} = \\ = \Pr\{\mathbf{e} \in \mathbf{h}^\perp\} = \sum_{i=0}^{n} A_i^\perp p^i (1-(q-1)p)^{n-i} = A_{\mathbf{h}^\perp}(p, 1-(q-1)p),\end{gathered} \tag{7}$$

где $A_{\mathbf{h}^\perp}(x, y) = \sum\limits_{i=0}^{n} A_i^\perp x^i y^{n-i}$ — энумератор весов кода $\mathbf{h}^\perp$; $A_i^\perp$ — число векторов веса i в коде $\mathbf{h}^\perp$. Для $[n, 1]$-кода $\langle \mathbf{h} \rangle$, порождённого вектором $\mathbf{h}$, рассмотрим энумератор весов $A_{\mathbf{h}}(x, y)$. По определению $A_{\mathbf{h}}(x, y) = \sum\limits_{i=0}^{n} A_i x^i y^{n-i}$, где A_i — число векторов веса i

в коде $\langle \mathbf{h} \rangle$. Так как $\dim(\langle \mathbf{h} \rangle) = 1$, то $A_{\mathbf{h}}(x, y) = y^n + (q-1)x^{\mathrm{w}(\mathbf{h})}y^{n-\mathrm{w}(\mathbf{h})}$. Воспользуемся тождеством Мак-Вильямс и получим

$$A_{\mathbf{h}^\perp}(x, y) = \frac{1}{q}((y + (q-1)x)^n + (q-1)(y-x)^{\mathrm{w}(\mathbf{h})}(y + (q-1)x)^{n-\mathrm{w}(\mathbf{h})}.$$

С учётом (7), подставив $x = p$ и $y = 1 - (q-1)p$ в $A_{\mathbf{h}^\perp}(x, y)$, получим (6). ■

Лемма 3. Пусть $\mathbf{z}$ — случайно и равновероятно выбранный из $\mathbb{F}_q^n$ вектор. Тогда $\Pr\{(\mathbf{z}, \mathbf{h}) = 0\} = q^{-1}$.

Доказательство. $\Pr\{(\mathbf{z}, \mathbf{h}) = 0\} = \Pr\{\mathbf{z} \in \mathbf{h}^\perp\}$. Но $\Pr\{\mathbf{z} \in \mathbf{h}^\perp\} = \dfrac{|\mathbf{h}^\perp|}{|\mathbb{F}_q^n|} = q^{-1}$. ■

Лемма 4. Пусть $\mathbf{c}$ — вектор, случайно и равновероятно выбранный из кода C' длины n; $C' \not\subseteq C$; $\mathbf{z} = \mathbf{c} + \mathbf{e}$ — вектор на выходе канала $q\mathrm{BC}(p)$. Тогда в любом базисе кода $C^\perp$ найдётся такой $\mathbf{h}$, что $\Pr\{(\mathbf{z}, \mathbf{h}) = 0\} = q^{-1}$.

Доказательство. Так как $C' \not\subseteq C$, в любом базисе кода $C^\perp$ всегда найдётся такой вектор $\mathbf{h}$, что $\mathbf{h} \notin C'^\perp$. Выберем такой $\mathbf{h}$. Пусть $\mathbf{h}_{\mathbf{a}}^\perp$ — это смежный класс подпространства $\mathbf{h}^\perp$ вида $\mathbf{h}_{\mathbf{a}}^\perp = \mathbf{a} + \mathbf{h}^\perp$, $\mathbf{a} \in \mathbb{F}_q^n$. Так как $\dim(\mathbf{h}^\perp) = n - 1$, то $|\mathbb{F}_q^n / \mathbf{h}^\perp| = q$. Рассмотрим множество $\mathcal{R} \subset \mathbb{F}_q^n$ мощности $q-1$, такое, что $\left(\bigcup\limits_{\mathbf{a}\in\mathcal{R}} \mathbf{h}_{\mathbf{a}}^\perp\right) \cup \mathbf{h}^\perp = \mathbb{F}_q^n$. Символом $-\mathbf{a}$ обозначим такой вектор, что $-\mathbf{a} + \mathbf{a} = \mathbf{0} \in \mathbb{F}_q^n$. Так как $-\mathbf{a} \in \mathcal{R}$ для любого $\mathbf{a} \in \mathcal{R}$, то

$$\begin{gathered}
\Pr\{(\mathbf{z}, \mathbf{h}) = 0\} = \Pr\{\mathbf{z} \in \mathbf{h}^\perp\} = \Pr\{\mathbf{c} + \mathbf{e} \in \mathbf{h}^\perp\} = \\
= \Pr\{\mathbf{c} \in \mathbf{h}^\perp\} \Pr\{\mathbf{e} \in \mathbf{h}^\perp\} + \sum_{\mathbf{a}\in\mathcal{R}} \Pr\{\mathbf{c} \in \mathbf{h}_{\mathbf{a}}^\perp\} \Pr\{\mathbf{e} \in \mathbf{h}_{-\mathbf{a}}^\perp\} = \\
= \Pr\{\mathbf{c} \in C' \cap \mathbf{h}^\perp\} \Pr\{\mathbf{e} \in \mathbf{h}^\perp\} + \sum_{\mathbf{a}\in\mathcal{R}} \Pr\{\mathbf{c} \in C' \cap \mathbf{h}_{\mathbf{a}}^\perp\} \Pr\{\mathbf{e} \in \mathbf{h}_{-\mathbf{a}}^\perp\} = \\
= \frac{|\mathbf{h}^\perp \cap C'|}{|C'|} \Pr\{\mathbf{e} \in \mathbf{h}^\perp\} + \sum_{\mathbf{a}\in\mathcal{R}} \frac{|\mathbf{h}_{\mathbf{a}}^\perp \cap C'|}{|C'|} \Pr\{\mathbf{e} \in \mathbf{h}_{-\mathbf{a}}^\perp\}.
\end{gathered}$$

Учитывая равенство $|\mathbf{h}_{\mathbf{a}}^\perp \cap C'| = |\mathbf{h}^\perp \cap C'|$ для всех $\mathbf{a} \in \mathcal{R}$, получим

$$\Pr\{(\mathbf{z}, \mathbf{h}) = 0\} = \frac{|\mathbf{h}^\perp \cap C'|}{|C'|} \left(\Pr\{\mathbf{e} \in \mathbf{h}^\perp\} + \sum_{\mathbf{a}\in\mathcal{R}} \Pr\{\mathbf{e} \in \mathbf{h}_{-\mathbf{a}}^\perp\}\right) = \frac{|\mathbf{h}^\perp \cap C'|}{|C'|}.$$

Пусть размерность кода C' равна r. Так как $\dim(\mathbf{h}^\perp) = n-1$ и $C' \not\subseteq \mathbf{h}^\perp$ (по построению $\mathbf{h} \notin C'^\perp$), то $\dim(\mathbf{h}^\perp \cap C') = r - 1$. Тогда $\dfrac{|\mathbf{h}^\perp \cap C'|}{|C'|} = \dfrac{q^{r-1}}{q^r} = \dfrac{1}{q}$. Отсюда следует доказываемое утверждение. ■

Пусть $\mathbf{h} \in \mathbb{F}_q^n$, $\mathbf{z} \in \mathbb{F}_q^n$. Рассмотрим случайную величину $X_{\mathbf{h},\mathbf{z}}$, принимающую значение 0, если $(\mathbf{h}, \mathbf{z}) = 0$, и значение 1 при $(\mathbf{h}, \mathbf{z}) \neq 0$. Если $\mathbf{c} \in \mathbf{h}^\perp$, $\mathbf{z} = \mathbf{c} + \mathbf{e}$ — вектор на выходе $q\mathrm{BC}(p)$, то соответствующую случайную величину $X_{\mathbf{h},\mathbf{z}}$ будем обозначать $X_{\mathbf{h},\mathbf{z}}^0$; если $\mathbf{c}' \in C' \not\subseteq \mathbf{h}^\perp$, $\mathbf{z} = \mathbf{c}' + \mathbf{e}$ — вектор на выходе $q\mathrm{BC}(p)$ или если $\mathbf{z}$ — случайно и равновероятно выбранный из $\mathbb{F}_q^n$ вектор, то случайную величину обозначим $X_{\mathbf{h},\mathbf{z}}^1$. Пусть $\mathbf{p}_0 = \Pr\{X_{\mathbf{h},\mathbf{z}}^0 = 1\}$, тогда из (6) получаем

$$\mathbf{p}_0 = 1 - \frac{1}{q}\left(1 + (q-1)(1-pq)^{\mathrm{w}(\mathbf{h})}\right) = \frac{q-1}{q}(1 - (1-pq)^{\mathrm{w}(\mathbf{h})}). \tag{8}$$

Пусть $\mathbf{p}_1 = \Pr\{X^1_{\mathbf{h},\mathbf{z}} = 1\}$. Из лемм 3 и 4 получаем

$$\mathbf{p}_1 = 1 - \frac{1}{q} = \frac{q-1}{q}. \tag{9}$$

Для $\mathbf{h} \in \mathbb{F}_q^n$, $\mathcal{S} \subseteq \mathbb{F}_q^n$ и $T \geqslant 0$ обозначим через $\mathrm{ST}(\mathbf{h}, \mathcal{S}, T)$ статистический критерий, согласно которому принимается решение о том, что выборка $\mathcal{S}$ является набором зашумлённых векторов из $\mathbf{h}^\perp$, если $\sum\limits_{\mathbf{z}\in\mathcal{S}} X_{\mathbf{h},\mathbf{z}} \leqslant T$. Рассмотрим случайные величины $X = \sum\limits_{\mathbf{z}\in\mathcal{S}}^{M} X_{\mathbf{h},\mathbf{z}}$ и $X_i = \sum\limits_{\mathbf{z}\in\mathcal{S}}^{M} X^i_{\mathbf{h},\mathbf{z}}$, $i = 0, 1$.

Теорема 1. Пусть $\mathbf{h} \in \mathbb{F}_q^n$, $\mathcal{S}$ — набор $M_{\mathbf{h}}$ векторов, полученный из канала $q\mathrm{SC}(p)$. При выборе статистического критерия $\mathrm{ST}(\mathbf{h}, \mathcal{S}, T_{\mathbf{h}})$ проверки гипотезы $H_0 : X = X_0$ против альтернативы $H_1 : X = X_1$ вероятности ошибок первого и второго рода не превышают α и β соответственно для

$$M_{\mathbf{h}} = \left(\frac{b\sqrt{(q-1)(1-(1-pq)^{\mathrm{w}(\mathbf{h})})(1+(q-1)(1-pq)^{\mathrm{w}(\mathbf{h})})} - a}{(\sqrt{q-1})(1-pq)^{\mathrm{w}(\mathbf{h})}} \right)^2, \tag{10}$$

$$T_{\mathbf{h}} = M_{\mathbf{h}} \frac{q-1}{q} + a\sqrt{M_{\mathbf{h}} \frac{q-1}{q^2}},$$

где $a = \Phi^{-1}(\alpha)$; $b = \Phi^{-1}(1-\beta)$; Φ^{-1} — обратная функция Лапласа.

Доказательство. Для каждого $i \in \{0, 1\}$ случайная величина X_i является суммой бернуллиевских случайных величин. Тогда по центральной предельной теореме случайная величина X_i асимптотически нормальна и её математическое ожидание равно $M_{\mathbf{h}}\mathbf{p}_i$, а дисперсия — $M_{\mathbf{h}}\mathbf{p}_i(1-\mathbf{p}_i)$. Пусть α и β — выбранные ошибки первого и второго рода соответственно при использовании статистического критерия $\mathrm{ST}(\mathbf{h}, \mathcal{S}, T_{\mathbf{h}})$. Тогда при проверке гипотезы $H_0 : X = X_0$ против альтернативы $H_1 : X = X_1$ имеют место равенства

$$\Pr\left\{\sum_{\mathbf{z}\in\mathcal{S}} X^1_{\mathbf{h},\mathbf{z}} \leqslant T_{\mathbf{h}}\right\} = \alpha, \quad \Pr\left\{\sum_{\mathbf{z}\in\mathcal{S}} X^0_{\mathbf{h},\mathbf{z}} \leqslant T_{\mathbf{h}}\right\} = 1-\beta.$$

Учитывая асимптотическую нормальность случайных величин X_0 и X_1, получим

$$\begin{cases} \Phi\left(\dfrac{T_{\mathbf{h}} - M_{\mathbf{h}}\mathbf{p}_1}{\sqrt{M_{\mathbf{h}}\mathbf{p}_1(1-\mathbf{p}_1)}}\right) = \alpha, \\ \Phi\left(\dfrac{T_{\mathbf{h}} - M_{\mathbf{h}}\mathbf{p}_0}{\sqrt{M_{\mathbf{h}}\mathbf{p}_0(1-\mathbf{p}_0)}}\right) = 1-\beta, \end{cases}$$

или

$$\begin{cases} T_{\mathbf{h}} = M_{\mathbf{h}}\mathbf{p}_1 + \Phi^{-1}(\alpha)\sqrt{M_{\mathbf{h}}\mathbf{p}_1(1-\mathbf{p}_1)}, \\ T_{\mathbf{h}} = M_{\mathbf{h}}\mathbf{p}_0 + \Phi^{-1}(1-\beta)\sqrt{M_{\mathbf{h}}\mathbf{p}_0(1-\mathbf{p}_0)}. \end{cases}$$

В результате имеем

$$\begin{cases} T_{\mathbf{h}} = M_{\mathbf{h}}\mathbf{p}_1 + \Phi^{-1}(\alpha)\sqrt{M_{\mathbf{h}}\mathbf{p}_1(1-\mathbf{p}_1)}, \\ M_{\mathbf{h}} = \left(\dfrac{\Phi^{-1}(1-\beta)\sqrt{\mathbf{p}_0(1-\mathbf{p}_0)} - \Phi^{-1}(\alpha)\sqrt{\mathbf{p}_1(1-\mathbf{p}_1)}}{\mathbf{p}_1 - \mathbf{p}_0}\right)^2. \end{cases}$$

Подставив значение $\mathbf{p}_0$ и $\mathbf{p}_1$ из (8) и (9), получим искомые выражения для $T_{\mathbf{h}}$ и $M_{\mathbf{h}}$. ■

Замечание 1. В доказательстве теоремы используется аппроксимация функции распределения F случайной величины X_i, $i \in \{0,1\}$, к функции Лапласа Ф. Поэтому в полученных формулах возникают погрешности, влияющие на доверительные вероятности. В соответствии с неравенством Берри — Эссеена [10, с. 75] погрешность для α меньше, чем $\dfrac{\mathbf{p}_0^2 + (1-\mathbf{p}_0)^2}{\sqrt{\mathbf{p}_0(1-\mathbf{p}_0)M_{\mathbf{h}}}}$, а для β меньше, чем $\dfrac{\mathbf{p}_1^2 + (1-\mathbf{p}_1)^2}{\sqrt{\mathbf{p}_1(1-\mathbf{p}_1)M_{\mathbf{h}}}}$.

Для каждого $p \in (0, 1/q)$ и $q \geqslant 2$ рассмотрим непрерывную функцию

$$A(x) = 1 + (q-2)(1-pq)^x - (q-1)(1-pq)^{2x}.$$

По построению, функция $A(x)$ для натуральных x всегда положительная. Непосредственно подставив в (10) выражение для $\mathbf{p}_1$ из формулы (9) и проведя необходимые преобразования, получим

$$M_{\mathbf{h}} = \left(\frac{b\sqrt{A(\mathrm{w}(\mathbf{h}))} - a}{\sqrt{q-1}(1-pq)^{\mathrm{w}(\mathbf{h})}} \right)^2.$$

Рассмотрим функцию $f(x) = \dfrac{bA(x)^{1/2} - a}{\sqrt{q-1}(1-pq)^x}$. Производная $f'(x)$ имеет вид

$$f'(x) = \frac{b(A(x)^{1/2})'\sqrt{q-1}(1-pq)^x - (bA(x)^{1/2} - a)\sqrt{q-1}(1-pq)^x \ln(1-pq)}{(q-1)(1-pq)^{2x}} =$$
$$= \frac{b(A(x)^{1/2})' - (bA(x)^{1/2} - a)\ln(1-pq)}{\sqrt{q-1}(1-pq)^x} =$$
$$= \frac{(-b/2A(x)^{1/2} - b/2A(x)^{-1/2} - b/2A(x)^{-1/2}(q-1)(1-pq)^{2x} + a)\ln(1-pq)}{\sqrt{q-1}(1-pq)^x}.$$

Отсюда получаем, что при $p \in (0, 1/q)$ и $q \geqslant 2$ производная равна нулю ($f'(x) = 0$) тогда и только тогда, когда

$$A(x)^{1/2} + A(x)^{-1/2} + A(x)^{-1/2}(q-1)(1-pq)^{2x} = \frac{2a}{b}. \tag{11}$$

Заметим, что $A(x)^{1/2} + A(x)^{-1/2}(1 + (q-1)(1-pq)^{2x}) \geqslant A(x)^{1/2} + A(x)^{-1/2} \geqslant 2$, так как функция $A(x)$ положительная для натуральных x и $y + 1/y \geqslant 2$ для $y > 0$. Если $a/b < 1$, то уравнение (11) решений не имеет, поэтому функция $f(x)$ (а также и $f^2(x)$) либо монотонно убывает, либо монотонно возрастает для $x > 0$. Для произвольных $p \in (0, 1/q)$ и $q \geqslant 2$ вычисления показывают, что $f(1) \leqslant f(2)$, следовательно, функция $f(x)$ возрастающая. Таким образом, справедлива

Лемма 5. Пусть $\mathbf{h} \neq \mathbf{h}'$, $w(\mathbf{h}) > w(\mathbf{h}')$ и в рамках условий теоремы 1 выполняется неравенство $a/b < 1$. Тогда $M_{\mathbf{h}} \geqslant M_{\mathbf{h}'}$.

Таким образом, чем меньше вес вектора $\mathbf{h}$, тем меньше может быть мощность выборки $\mathcal{S}$ зашумлённых векторов для применения статистического критерия $\mathrm{ST}(\mathbf{h}, \mathcal{S}, T_{\mathbf{h}})$.

Ниже приведён алгоритм 1 проверки гипотезы H_0 против альтернативы H_1. Естественно считать, что чем меньший размер выборки требуется для применения статистического критерия при заданных уровнях ошибок первого и второго рода, тем эффективнее критерий. В соответствии с леммой 5, для применения статистического критерия следует выбирать базисные векторы малого веса. Для этого нам понадобится

понятие минимального базиса кода, а именно: минимальным базисом кода $\mathcal{C}$ будем называть такой базис $B^0_{\mathcal{C}} \in \mathcal{B}_{\mathcal{C}}$, что в любом другом базисе $B_{\mathcal{C}} \in \mathcal{B}_{\mathcal{C}}$ найдётся базисный вектор, вес которого не меньше веса любого вектора из $B^0_{\mathcal{C}}$. Отметим, что минимальный базис определяется неоднозначно.

Алгоритм 1. CodeRecognition$(\mathcal{C}, \mathcal{S}, p, \alpha, \beta)$ (распознавание кода)

Вход: $\mathcal{C}$ — предполагаемый код, $\mathcal{S}$ — выборка, p — вероятность ошибки в $q\mathrm{SC}(p)$, α и β — доверительные вероятности ошибок 1 и 2 рода для (10).

1: Найти минимальный базис $B^0_{\mathcal{C}^\perp}$ кода $\mathcal{C}^\perp$.
2: **Для всех** $\mathbf{h} \in B^0_{\mathcal{C}^\perp}$
3: найти $M_{\mathbf{h}}$ и $T_{\mathbf{h}}$ согласно теореме 1;
4: выбрать набор $\mathcal{S}_{\mathbf{h}}$ из $\mathcal{S}$ мощности $M_{\mathbf{h}}$;
5: проверить с помощью критерия $\mathrm{ST}(\mathbf{h}, \mathcal{S}_{\mathbf{h}}, T_{\mathbf{h}})$ принадлежность выборки $\mathcal{S}_{\mathbf{h}}$ каждому пространству $\mathbf{h}^\perp$.
6: **Если** выборка $\mathcal{S}_{\mathbf{h}}$ является набором зашумлённых векторов из $\mathbf{h}^\perp$ для каждого $\mathbf{h}$ из $B^0_{\mathcal{C}^\perp}$, **то**
7: выборка $\mathcal{S}$ полагается набором зашумлённых кодовых слов кода $\mathcal{C}$, в противном случае считается, что $\mathcal{S}$ не является набором зашумлённых кодовых слов кода $\mathcal{C}$.

На шаге 1 алгоритма предлагается искать минимальный базис, так как в этом случае потребуется наименьшее в рамках предлагаемого способа количество наблюдаемых векторов. Для произвольного кода задача нахождения минимального базиса представляется трудной, но для некоторых кодов такой базис может быть найден легко. К последним относятся коды с максимальным достижимым расстоянием (МДР-коды) и равновесные коды. Так как порождающая матрица линейного $[n, n-k]$-кода $\mathcal{C}^\perp$ может быть приведена к псевдосистематическому виду [7], то вес каждого вектора в $B^0_{\mathcal{C}^\perp}$ не превышает $k+1$. Из лемм 2 и 5 следует

Теорема 2. Пусть $\mathcal{C}$ — $[n, k]$-код, $M(\mathcal{C}, \alpha, \beta)$ — минимально необходимое количество векторов, достаточное для проверки алгоритмом CodeRecognition гипотезы H_0 против гипотезы H_1. Тогда $M(\mathcal{C}, \alpha, \beta) \leqslant f^2(k+1)$, причём равенство выполняется в случае, когда $\mathcal{C}$ — МДР-код.

3. Поиск подвектора в подслушанных данных

Рассмотрим ИАС, описанную в п. 1. Предлагается следующий способ (алгоритм 2) определения длины и позиции сообщения $\mathbf{m}$, основанный на лемме 1. Он заключается в том, что если выборка $\widetilde{\mathbf{Z}}$ вида (4) для гипотезы $H_{i,j}$ представляет собой набор кодовых слов кода $\mathcal{C}^{(i,j)}$, то гипотеза $H_{i,j}$ верна. Тот факт, что выборка (4) является набором зашумлённых кодовых слов кода $\mathcal{C}^{(i,j)}$, можно проверить с помощью статистического критерия $\mathrm{ST}(\mathbf{h}, \widetilde{\mathbf{Z}}, T)$, где $\mathbf{h}$ пробегает базис кода $(\mathcal{C}^{(i,j)})^\perp$, а объём выборки $\widetilde{\mathbf{Z}}$ и параметр T для каждого $\mathbf{h}$ определяются в соответствии с теоремой 1. Таким образом, проверяя гипотезы, можно найти длину m_l и позицию m_s сообщения $\mathbf{m}$. Заметим, что при проверке текущей гипотезы $H_{i,j}$ необязательно перехватывать новые сообщения, так как можно использовать уже перехваченные сообщения, использовавшиеся для проверки предыдущих гипотез. Если же для какой-либо гипотезы недостаточно перехваченных ранее сообщений, то необходимо дополнительно перехватить ровно столько, сколько не хватило.

Поясним алгоритм ShiftAndLenFinding. Сначала ищется начальная координата неизвестного вектора $\mathbf{m}$, при этом на основании п. 1 леммы 1 полагаем, что длина этого

Алгоритм 2. ShiftAndLenFinding($\mathcal{S}, p, \alpha, \beta, P$) (нахождение длины и позиции)

Вход: $\mathcal{S}$ — выборка, p — вероятность ошибки в qSC(p), α и β — доверительные вероятности ошибок 1 и 2 рода для (10), P — матрица из (1).

1: $i := 1$, $j := 1$, $m_s := -1$.
2: В рамках гипотезы $H_{i,j}$ построить матрицу $[P_2^{(i,j)}|I_{n-k}]$ и найти $\widehat{p}$ (см. лемму 1).
3: Для кода $\mathcal{C}^{(i,j)}$ с порождающей матрицей $G(\mathcal{C}^{(i,j)}) = [P_2^{(i,j)}|I_{n-k}]$ вычислить $M = \max\{M_{\mathbf{h}} : \mathbf{h} \in B^0_{(\mathcal{C}^{(i,j)})^\perp}\}$ в соответствии с (10) для $p = \widehat{p}$.
4: **Если** $|\mathcal{S}| < 2M$, **то** перехватить недостающее количество векторов.
5: Построить выборку $\mathcal{S}^{(i,j)}$ вида (4).
6: Проверить гипотезу с помощью алгоритма CodeRecognition($\mathcal{C}^{(i,j)}, \mathcal{S}^{(i,j)}, \widehat{p}, \alpha, \beta$).
7: **Если** $m_s > 0$, **то** перейти на шаг 9.
8: **Если** гипотеза неверна, **то** $i := i + 1$ и возврат к шагу 2, **иначе** $m_s := i$.
9: **Если** гипотеза верна, **то** $j := j + 1$ и возврат к шагу 2, **иначе** $m_l := j - 1$.

Выход: m_s, m_l — позиция и длина сообщения $\mathbf{m}$.

вектора равна 1. На этапе нахождения начальной координаты гипотеза H_0 проверяется против альтернативы H_1, так как при ошибочном выборе начальной координаты (и при предполагаемой единичной длине секретного подвектора) набор (4) представляет собой набор случайных и равновероятных векторов (при условии, что ключи $\mathbf{K}^{(t)}$ и векторы $\mathbf{m}_i^{(t)}$, $i \in \{1, 2\}$, случайны и равновероятны, см. п. 1). Когда предполагаемая позиция вектора $\mathbf{m}$ найдена, начинается этап поиска длины путём постепенного увеличения, начиная с единицы. При этом если длина будет превышать истинную длину вектора $\mathbf{m}$, то набор (4) уже будет представлять собой зашумлённые векторы некоторого кода, содержащего код $\mathcal{C}^{(i,j)}$. В частности, если $i = m_s$ и $j = m_l + 1$ (то есть позиция вектора угадана верно, но предполагаемая длина на единицу больше, чем истинная), построенная выборка (4) будет представлять собой векторы линейного $[n-k+j, n-k]$-кода, порождённого базисом кода $\mathcal{C}^{(m_s, m_l+1)}$ и вектором $(\mathbf{0}^1 \parallel 1 \parallel \mathbf{0}^2)$, где $\mathbf{0}^1$ и $\mathbf{0}^2$ — нулевые векторы длины m_l и $n - k$ соответственно. Но так как вектор $\mathbf{m}_2^{(t)}$ случайный и равновероятный, каждый вектор в наборе (4) с вероятностью $1 - 1/q^2$ принадлежит смежному классу $(\mathbf{0}^1 \parallel 1 \parallel \mathbf{0}^2) + \mathcal{C}^{(m_s, m_l+1)}$ кода $\mathcal{C}^{(m_s, m_l+1)}$. При правильно найденных длине m_l и позиции m_s выборка $\mathbf{Z}_{\tau(m_s, m_l)}$ вида (3) представляет собой набор векторов вида

$$\widehat{\mathbf{z}}^{(t)} = (\mathbf{m} + \mathbf{K}^{(t)} P_2^{(m_s, m_l)} \parallel \mathbf{K}^{(t)}) + (\mathbf{e}_2^{(t)} \parallel \mathbf{e}_4^{(t)}), \quad t = 1, 2, \ldots \tag{12}$$

Тогда задача восстановления сообщения $\mathbf{m}$ может быть решена, например, полным перебором по всевозможным значениям из $\mathbb{F}_q^{m_l}$ с помощью алгоритма CodeRecognition (если размер поля q и длина m_l сообщения позволяют произвести полный перебор за приемлемое время). Соответствующий алгоритм 3 построен ниже.

Так как $\widehat{p} \geqslant p$ (п. 1 леммы 1), для применения алгоритма ShiftAndLenFinding потребуется как минимум в 2 раза больше перехватов, чем для применения алгоритма SubvectorFinding. Таким образом, при нахождении содержания сообщения может быть использована выборка, построенная во время выполнения ShiftAndLenFinding. Для матрицы P (см. правило (1)) обозначим $M(P)$ минимальное число векторов, достаточное для нахождения алгоритмами ShiftAndLenFinding и SubvectorFinding вектора $\mathbf{m}$.

Алгоритм 3. SubvectorFinding($\mathcal{S}, p, \alpha, \beta$) (нахождение значения вектора)

Вход: $\mathcal{S} = \{\mathbf{z}^{(t)}\}_{t=1,2,\ldots}$ — выборка вида (3) при $i = m_s$ и $j = m_l$, p — вероятность ошибки в qSC(p), α и β — вероятности ошибок 1 и 2 рода для (10).

1: $\mathcal{M} := \varnothing$.
2: **Для всех** $\mathbf{y} \in \mathbb{F}_q^{m_l}$
3: по выборке $\mathcal{S}$ построить выборку $\widehat{\mathcal{S}} = \{\widehat{\mathbf{z}}^{(t)}\}_{t=1,2,\ldots}$, где $\widehat{\mathbf{z}}^{(t)} = \mathbf{z}^{(t)} - (\mathbf{y} \parallel \mathbf{0})$, $\mathbf{0}$ — нулевой вектор пространства $\mathbb{F}_q^{n-k}$;
4: проверить с помощью алгоритма CodeRecognition($\mathcal{C}^{(m_s,m_l)}, \widehat{\mathcal{S}}, p, \alpha, \beta$) гипотезу о том, что выборка $\widehat{\mathcal{S}}$ порождена кодовыми словами кода $\mathcal{C}^{(m_s,m_l)}$.
5: **Если** гипотеза верна, **то** $\mathcal{M} := \mathcal{M} \cup \{\mathbf{y}\}$.

Выход: $\mathcal{M}$.

Замечание 2. Если $\mathbf{m} \neq \mathbf{y}$, но $(\mathbf{m} - \mathbf{y} \parallel \mathbf{0}) \in \mathcal{C}^{(m_s,m_l)}$, то алгоритм SubvectorFinding добавит вектор $\mathbf{y}$ в множество кандидатов $\mathcal{M}$. Так как порождающая матрица $[n - k + m_l, n - k]$-кода $\mathcal{C}^{(m_s,m_l)}$ имеет вид $G(\mathcal{C}^{(m_s,m_l)}) = [P_2^{(m_s,m_l)}|I_{n-k}]$, то $(\mathbf{m} - \mathbf{y} \parallel \mathbf{0}) \in \mathcal{C}^{(m_s,m_l)}$ только в случае $\mathbf{m} = \mathbf{y}$. Поэтому существуют такие значения α и β, при которых алгоритм SubvectorFinding возвращает одноэлементное множество кандидатов.

Отметим, что нахождение сообщения $\mathbf{m}$ по выборке $\mathbf{Z}_{\tau(m_s,m_l)}$, состоящей из векторов вида (12), возможно и неперебо́рным методом. Один из таких методов может быть построен для двоичного случая ($q = 2$) на основе результатов работы [3].

Для примера рассмотрим $[n, n-k]$-код $\mathcal{C}$ Рида — Соломона над полем $\mathbb{F}_q$, $q > 2$. Пусть порождающая матрица кода $\mathcal{C}$ имеет вид $G(\mathcal{C}) = [P|I_{n-k}]$. Заметим, что код $\mathcal{C}$ — это МДР-код, поэтому для любого $\omega \subset \{1, \ldots, k\}$ и $\tau = \omega \cup \{k+1, \ldots, n\}$ код $\pi_\tau(\mathcal{C})$ с порождающей матрицей $\pi_\tau(G(\mathcal{C})) = [\pi_\omega(P)|I_{n-k}]$ является МДР-кодом размерности $n-k$ и длины $|\omega|+n-k$. Так как $\pi_\tau(G(\mathcal{C}))^\perp$ — также МДР-код, его кодовое расстояние равно $n-k+1$ (не зависит от ω и его мощности). Пусть в правиле (1) используется матрица P, являющаяся частью порождающей матрицы $G(\mathcal{C})$ рассматриваемого кода Рида — Соломона $\mathcal{C}$. Тогда построенная на шаге 2 алгоритма ShiftAndLenFinding матрица $\pi_{\tau(i,j)}(P) = [P_2^{(i,j)}|I_{n-k}]$ является матрицей МДР-кода $\pi_{\tau(i,j)}(\mathcal{C})$ размерности $n-k+1$ и по теореме 2 минимальный базис $B^0_{\pi_{\tau(i,j)}(\mathcal{C})^\perp}$, строящийся на первом шаге алгоритма CodeRecognition (к этому алгоритму обращение происходит на шаге 6 алгоритма определения позиции и длины постоянного подвектора), состоит из векторов веса $n-k+1$. Отметим, что базис $B^0_{\pi_{\tau(i,j)}(\mathcal{C})^\perp}$, например, может представлять собой строки матрицы $[I_j|(-P_2^{(i,j)})^\top]$ — порождающей матрицы кода $\pi_{\tau(i,j)}(\mathcal{C})^\perp$, где $A^\top$ — транспонированная матрица A. Таким образом, размер выборки для определения позиции и длины сообщения, а также содержания сообщения определяется на основании длины ключа, используемого в правиле (1). Результаты вычислений минимального количества перехваченных векторов для нахождения подвектора $\mathbf{m}$ в зашумленных данных приведены в таблице для $q = 16$, $n = 15$, $\alpha = 0{,}05$, $\beta = 0{,}04$.

Значение величны $M(P)$ в случае, когда $[P|I_{n-k}]$ — порождающая матрица кода Рида — Соломона, $q = 16$, $n = 15$, $\alpha = 0{,}05$, $\beta = 0{,}04$

	p							
$n-k$	$\leqslant 0{,}0125$	0,01875	0,025	0,03125	0,0375	0,04375	0,05	0,05625
1	2	2	2	2	2	2	2	2
2	2	2	2	2	2	2	2	213
3	2	2	2	2	2	5	519	2,0 E+6
4	2	2	2	2	6	494	3,2 E+5	2,1 E+10
5	2	2	2	3	194	60604	2,0 E+8	2,1 E+14
6	2	2	2	37	7515	7,5 E+6	1,3 E+11	2,1 E+18
7	2	2	4	570	293173	9,2 E+8	7,9 E+13	2,1 E+22
8	2	2	28	9083	1,1 E+7	1,1 E+11	4,9 E+16	2,1 E+26
9	2	2	207	145186	4,5 E+8	1,4 E+13	3,1 E+19	2,1 E+30
10	2	4	1585	2,3 E+6	1,7 E+10	1,7 E+15	1,9 E+22	2,1 E+34
11	2	15	12205	3,7 E+7	6,8 E+11	2,1 E+17	1,2 E+25	2,1 E+38
12	2	59	94107	5,9 E+8	2,7 E+13	2,6 E+19	7,5 E+27	2,1 E+42
13	2	242	725943	9,5 E+9	1,0 E+15	3,3 E+21	4,7 E+30	2,1 E+46
14	2	1003	5,6 E+6	1,5 E+11	4,1 E+16	4,0 E+23	2,9 E+33	2,1 E+50

Теорема 3. Пусть q — степень простого числа; $q\mathrm{SC}(p)$ — q-ичный симметричный канал с вероятностью ошибки p; $n, k \in \mathbb{N}$, $k < n$; $\alpha, \beta \in [0; 1]$; P и $\widetilde{P}$ — разные $((n-k) \times k)$-матрицы, причём матрица $[P|I_{n-k}]$ является порождающей матрицей МДР-кода. Тогда $M(P) \geqslant M(\widetilde{P})$.

Доказательство. Ранее показано, что для МДР-кода минимальный базис дуального кода состоит из векторов веса $n-k+1$. Заметим, что для не-МДР-кода минимальный базис дуального кода состоит из векторов веса, не превышающего $n-k+1$. На основании теоремы 2 получаем доказываемое утверждение. ■

Таким образом, при заданных q, p, n, k, α и β лучшую стойкость метода кодового зашумления к атаке нахождения фиксированного вектора обеспечивают такие $((n-k) \times k)$-матрицы P, для которых матрица $[P|I_{n-k}]$ порождает МДР-код, так как в этом случае потребуется наибольшее количество перехватов для нахождения сообщения.

ЛИТЕРАТУРА

1. *Wyner A. D.* The wire-tap channel // Bell Sys. Tech. J. 1975. V. 54. P. 1355–1387.
2. *Коржик В. И., Яковлев В. А.* Неасимптотические оценки эффективности кодового зашумления одного канала // Пробл. передачи информ. 1981. Т. 17. № 4 С. 11–18.
3. *Иванов В. А.* Статистические методы оценки эффективности кодового зашумления // Труды по дискретной математике. Т. 6. М.: Физматлит, 2002. С. 48–63.
4. *Косолапов Ю. В., Турченко О. Ю.* Поиск информационного сообщения в зашумлённых кодовых блоках при многократной передаче данных // Прикладная дискретная математика. Приложение. 2016. № 9. С. 55–57.
5. *Weidmann C.* Coding for the q-ary symmetric channel with moderate q // IEEE Int. Symp. Inform. Theory. 2008. P. 2156–2159.
6. *Couvreur A.* Distinguisher-based attacks on public-key cryptosystems using Reed — Solomon codes // Designs, Codes and Cryptography. 2014. V. 73. No. 2. P. 641–666.
7. *Сидельников В. М.* Теория кодирования. М.: Физматлит, 2008. 324 с.
8. *Chabot C.* Recognition of a code in a noisy environment // Proc. IEEE ISIT. June 2007. P. 2211–2215.

9. *Yardi A. D. and Vijayakumaran S.* Detecting linear block codes in noise using the GLRT // Proc. IEEE Intern. Conf. Communications, Budapest, Hungary, June 9–13, 2013. P. 4895–4899.
10. *Ширяев А. Н.* Вероятность. В 2-х кн. 3-е изд., перераб. и доп. М.: МЦНМО, 2004. Кн. 1 — 520 с., кн. 2 — 408 с.

REFERENCES

1. *Wyner A. D.* The wire-tap channel. Bell Sys. Tech. J., 1975, vol. 54, pp. 1355–1387.
2. *Korzhik V. I. and Yakovlev V. A.* Neasimptoticheskie otsenki effektivnosti kodovogo zashumleniya odnogo kanala [Nonasymptotic estimates for efficiency of code jamming in a wire-tap channel]. Probl. Peredachi Inf., 1981, vol. 17, iss. 4, pp. 11–18. (in Russian)
3. *Ivanov V. A.* Statisticheskie metody otsenki effektivnosti kodovogo zashumleniya [Statistical estimation of the code noising efficiency]. Tr. Diskr. Mat., 2002, vol. 6, pp. 48–63. (in Russian)
4. *Kosolapov Y. V. and Turchenko O. Y.* Poisk informatsionnogo soobshcheniya v zashumlennykh kodovykh blokakh pri mnogokratnoy peredache dannykh [Search of an information message in noisy code blocks at repeated data transmission]. Prikladnaya Diskretnaya Matematika. Prilozhenie, 2016, no. 9, pp. 55–57. (in Russian)
5. *Weidmann C.* Coding for the q-ary symmetric channel with moderate q. IEEE Int. Symp. Inform. Theory, 2008, pp. 2156–2159.
6. *Couvreur A.* Distinguisher-based attacks on public-key cryptosystems using Reed — Solomon codes. Designs, Codes and Cryptography, 2014, vol. 73, no. 2, pp. 641–666.
7. *Sidel'nikov V. M.* Teoriya Kodirovaniya [Coding Theory]. Moscow, Fizmatlit Publ., 2008, 324 p. (in Russian)
8. *Chabot C.* Recognition of a code in a noisy environment. Proc. IEEE ISIT, June 2007, pp. 2211–2215.
9. *Yardi A. D. and Vijayakumaran S.* Detecting linear block codes in noise using the GLRT. Proc. IEEE Intern. Conf. Communications, Budapest, Hungary, June 9–13, 2013, pp. 4895–4899.
10. *Shiryaev A. N.* Veroyatnost' [Probability]. Moscow, MCCME Publ., 2004. (in Russian)