УДК 517.19 DOI 10.17223/2226308X/9/22

ПОИСК ИНФОРМАЦИОННОГО СООБЩЕНИЯ В ЗАШУМЛЁННЫХ КОДОВЫХ БЛОКАХ ПРИ МНОГОКРАТНОЙ ПЕРЕДАЧЕ ДАННЫХ

Ю. В. Косолапов, О. Ю. Турченко

Рассматривается модель защиты данных с помощью метода кодового зашумления. Предполагается, что кодируемые информационные блоки длины k содержат фиксированное сообщение m длины $l \leqslant k$ на фиксированной позиции q $(1 \leqslant q \leqslant k - l + 1)$, а наблюдатель получает зашумлённые кодовые слова длины n через двоичный симметричный канал с вероятностью ошибки $(1 - \Delta)/2$, $0 < \Delta \leqslant 1$. Целью наблюдателя является нахождение неизвестного ему сообщения m, когда позиция q неизвестна, а длина l известна. Предложен способ нахождения сообщения m и получена оценка количества наблюдаемых кодовых слов, достаточного для восстановления сообщения m этим способом.

Ключевые слова: *кодовое зашумление, многократная передача данных.*

Рассмотрим информационно-аналитическую систему (ИАС), в которой два легальных участника (отправитель и получатель) связаны каналом без помех, а пассивный наблюдатель подслушивает передаваемые данные по двоичному симметричному каналу с вероятностью ошибки $(1 - \Delta)/2$, $0 < \Delta \leqslant 1$. Такая система впервые была исследована в [1]. Предполагается, что перед передачей в канал данные кодируются с помощью метода случайного кодирования, а именно: для зафиксированных натуральных чисел k и n $(k < n)$ легальными участниками выбрана $((n-k)\times k)$-матрица P с элементами из поля $\mathbb{F}_2$, а каждый k-битный вектор s кодируется по правилу

$$\mathrm{Enc}(s) = (c_1|K) = c, \tag{1}$$

где $c_1 = s \oplus KP$; K — случайно и равновероятно выбранный вектор из $\mathbb{F}_2^{n-k}$; запись $a|b$ обозначает конкатенацию векторов a и b. Предполагается, что 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 $\mathrm{Dec}(c) = KP \oplus c_1 = s$. У наблюдателя при подслушивании одного кодового слова из-за наличия помех возникает неопределённость относительно сообщения, которое было закодировано. Вычислению этой неопределённости посвящена, например, 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 сообщений, соответствующих одному информационному сообщению, достаточного для нахождения этого сообщения с заданными вероятностями ошибок первого и второго рода.

В настоящей работе рассматривается более сложная задача нахождения информационного сообщения в рамках модели многократной передачи данных, а именно предполагается, что в момент времени $t \in \mathbb{N}$ информационное сообщение $s^{(t)} \in \mathbb{F}_2^k$ имеет вид $s^{(t)} = (m_1^{(t)}|m|m_2^{(t)})$, $m_1^{(t)} \in \mathbb{F}_2^{q-1}$, $m \in \mathbb{F}_2^l$, $m_2^{(t)} \in \mathbb{F}_2^{k-[l+q]+1}$, где при $i \neq j$ в общем случае $\mathsf{P}\{m_1^{(i)} = m_1^{(j)}\} \neq 1$ и $\mathsf{P}\{m_2^{(i)} = m_2^{(j)}\} \neq 1$, а сообщение m постоянное для всех t. Предполагается также, что наблюдателю позиция q сообщения m неизвестна, а его длина l известна. Целью наблюдателя является нахождение неизвестного сообщения m при многократной передаче сообщений $s^{(t)}$, закодированных по правилу (1).

Задача решается в два этапа: сначала находится позиция q, а затем — само сообщение m. Для нахождения позиции q предлагается следующий способ. Выдвига-

ется гипотеза H_i о том, что $q = i$. В этом случае матрица P представима в виде $P = [P_1^{(i)}|P_2^{(i)}|P_3^{(i)}]$, где $P_1^{(i)}$ — первые $i-1$ столбцов матрицы P; $P_2^{(i)}$ — столбцы матрицы P с номерами от i до $i+l-1$; $P_3^{(i)}$ — последние $(k-(i+l)+1)$ столбцов матрицы P. В рамках гипотезы наблюдаемый в момент времени t вектор $z^{(t)} = c^{(t)} \oplus \theta^{(t)}$ имеет вид

$$z^{(t)} = (\widehat{m}_1^{(t)} \oplus K^{(t)} P_1^{(i)} \oplus \theta_1^{(t)} | \widehat{m} \oplus K^{(t)} P_2^{(i)} \oplus \theta_2^{(t)} | \widehat{m}_2^{(t)} \oplus K^{(t)} P_3^{(i)} \oplus \theta_3^{(t)} | K^{(t)} \oplus \theta_4^{(t)}),$$

где $\theta^{(t)} = (\theta_1^{(t)}|\theta_2^{(t)}|\theta_3^{(t)}|\theta_4^{(t)})$ — вектор помехи в двоичном симметричном канале.

Пусть $\tau(i) = \{i, \ldots, i+l-1\} \cup \{k+1, \ldots, n\}$ — множество координат. Построим выборку

$$Z_{\tau(i)} = \left(\widehat{z}_{\tau(i)}^{(1)}, \widehat{z}_{\tau(i)}^{(2)}, \ldots, \widehat{z}_{\tau(i)}^{(N)}\right), \tag{2}$$

где $\widehat{z}_{\tau(i)}^{(t)} = \pi_{\tau(i)}(z^{(t)})$ — проекция вектора $z^{(t)}$ на множество координат $\tau(i)$. Рассмотрим суммы $\widetilde{z}^{(t)} = \widehat{z}_{\tau(i)}^{(2t)} \oplus \widehat{z}_{\tau(i)}^{(2t-1)}$ и по ним сконструируем набор векторов:

$$\widetilde{Z} = \left(\widetilde{z}^{(1)} = \widehat{z}_{\tau(i)}^{(2)} \oplus \widehat{z}_{\tau(i)}^{(1)}, \widetilde{z}^{(2)} = \widehat{z}_{\tau(i)}^{(4)} \oplus \widehat{z}_{\tau(i)}^{(3)}, \ldots, \widetilde{z}^{(N/2)} = \widehat{z}_{\tau(i)}^{(N)} \oplus \widehat{z}_{\tau(i)}^{(N-1)}\right). \tag{3}$$

Утверждение 1. Если H_i — верная гипотеза ($i = q$), то выборка (3) представляет собой набор из зашумлённых кодовых слов линейного $[n-k+l, n-k]$-кода $\mathcal{D}^{(i)}$ с порождающей матрицей $[P_2^{(i)}|I_{n-k}]$, полученных из двоичного симметричного канала с вероятностью помехи $(1-\Delta^2)/2$.

В силу утверждения 1, проверка верности гипотезы H_i сводится к задаче распознавания кода по зашумлённому набору векторов. Заметим, что задача распознавания кода по зашумлённой выборке имеет несколько способов решения [4, 5]). В настоящей работе применяется метод из [4], идея которого состоит в том, что вес синдрома зашумлённого кодового слова в среднем меньше веса синдрома произвольного вектора. В соответствии с [4] для $h \in \mathbb{F}_2^v$, $\mathcal{Z} \subseteq \mathbb{F}_2^v$ и $T \geqslant 0$ обозначим через $\mathrm{ST}(h, \mathcal{Z}, T)$ статистический критерий, согласно которому принимается решение о том, что выборка $\mathcal{Z}$ является набором зашумлённых векторов из $h^\perp$, если $\sum\limits_{\widetilde{z} \in \mathcal{Z}} (h, \widetilde{z}) \leqslant T$.

Теорема 1 (C. Chabot, [4]). Пусть $h \in \mathbb{F}_2^v$, $w(h)$ — вес Хэмминга вектора h, $\mathcal{Z}$ — выборка из M векторов, полученная из двоичного симметричного канала с вероятностью ошибки p. При выборе статистического критерия $\mathrm{ST}(h, \mathcal{Z}, T)$ вероятности ошибок первого и второго рода не превышают α и β соответственно для

$$M = \left(\frac{b\sqrt{1-(1-2p)^{2w(h)}} - a}{(1-2p)^{w(h)}}\right)^2, \quad T = \frac{1}{2}(M + a\sqrt{M}), \tag{4}$$

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

Обозначим через $M(h, \alpha, \beta, p)$ объём выборки, вычисленный по формуле (4) для заданных h, α, β и p.

Утверждение 2. В рамках условий теоремы 1, для определения того, что выборка $\widetilde{Z}$ вида (3) является набором зашумлённых кодовых слов кода $\mathcal{D}^{(i)}$, размер $N/2$ выборки $\widetilde{Z}$ должен удовлетворять условию

$$N/2 \geqslant \max_{h \in \mathcal{H}_i} \{M(h, \alpha, \beta, (1-\Delta^2)/2)\},$$

где $\mathcal{H}_i$ — базис кода ${\mathcal{D}^{(i)}}^\perp$, состоящий из векторов малого веса.

Таким образом, для определения позиции q неизвестного сообщения m потребуется перехватить не менее $N_1 = 2\max\limits_i \max\limits_{h\in\mathcal{H}_i}\{M(h,\alpha,\beta,(1-\Delta^2)/2)\}$ зашумлённых кодовых сообщений, где $i \in \{1,\ldots,k-l+1\}$.

Пусть $q = i$ — определённая позиция сообщения m после наблюдения N_1 зашумлённых кодовых слов. Чтобы оценить необходимый объём выборки для восстановления сообщения m, используем метод восстановления фиксированного сообщения при кодовом зашумлении; метод описан в [3]. Если гипотеза H_i верная, то для применения этого метода нужно использовать выборку (2). Пусть N_2 — необходимое число перехватов для восстановления исходного сообщения m по выборке (2). Тогда минимальное число перехватов зашумлённых кодовых слов, необходимое для восстановления исходного сообщения m, равно $\max(N_1, N_2)$.

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. 2002. Т. 6. С. 48–63.
4. *Chabot C.* Recognition of a code in a noisy environment // Proc. IEEE ISIT, June 2007. P. 2211–2215.
5. *Yardi A. D. and Vijayakumaran S.* Detecting linear block codes in noise using the GLRT // Proc. IEEE Intern. Conf. Communications (ICC), Budapest, Hungary, June 9–13, 2013. P. 4895–4899.

УДК 519.6 DOI 10.17223/2226308X/9/23

О ТОЧНОСТИ МАТРИЧНО-ГРАФОВОГО ПОДХОДА К ОЦЕНКЕ ПЕРЕМЕШИВАЮЩИХ СВОЙСТВ ПРЕОБРАЗОВАНИЙ

С. Н. Кяжин, Ф. В. Лебедев

Приведены экспериментальные результаты оценки точности матрично-графового подхода к исследованию перемешивающих свойств нелинейных преобразований. В качестве класса преобразований, для которого проводилась оценка, взяты все преобразования множества V_n двоичных n-мерных векторов, перемешивающий граф которых есть n-вершинный граф Виландта, а также раундовые подстановки алгоритмов блочного шифрования AES, «Кузнечик» и «Магма» (ГОСТ 28147-89). Установлено, что полученные при матрично-графовом подходе оценки точны для 25 % преобразований с перемешивающим графом Виландта ($n = 9, 10, 11$), а также для раундовой подстановки алгоритмов AES и «Кузнечик». Указанные оценки не являются точными для раундовых подстановок алгоритма «Магма» и для 75 % преобразований с перемешивающим графом Виландта.

Ключевые слова: *перемешивающие свойства, матрично-графовый подход, граф Виландта, AES, «Кузнечик», «Магма».*

Введение

Для оценки перемешивающих свойств с помощью композиции преобразований используют оценочный матрично-графовый подход [1, гл. 10]. Начиная с 2010 г., в журнале «Прикладная дискретная математика» опубликован ряд теоретических и прикладных работ, развивающих данный подход (обзор результатов до 2012 г. см. в [2]).