О вычислении меры стойкости кодового зашумления в канале со случайным частичным перехватом

Ю. В. Косолапов[1], Н. О. Курчев[1]

Получена формула вычисления меры стойкости кодового зашумления для модели случайного перехвата. Для частных случаев, когда случайный перехват описывается моделями равновероятного случайного перехвата и перехвата, моделируемого симметричным каналом со стираниями, получены формулы для вычисления меры стойкости на основе спектра IDLP кодов.

Ключевые слова: случайный перехват, кодовое зашумление, наблюдатель, неопределённость.

Введение

В работе [1] рассмотрена модель передачи данных от абонента $\mathcal{S}$ к абоненту $\mathcal{D}$ в условиях присутствия пассивного наблюдателя $\mathcal{W}$. В ряде случаев процесс наблюдения можно представить в виде передачи данных от абонента $\mathcal{S}$ наблюдателю $\mathcal{W}$ по каналу, в общем случае отличному от канала, связывающего $\mathcal{S}$ и $\mathcal{D}$. Тогда модель передачи данных с наблюдением имеет вид, изображённый на рис. 1. В [1] доказано, что если источник $\mathcal{S}$ является случайным и равновероятным, а канал Σ_m, связывающий легальных абонентов и обычно называемый главным каналом, в определённом смысле лучше канала наблюдения Σ_w, то найдётся такой код с ненулевой скоростью передачи данных, что абонент $\mathcal{D}$ получит переданное сообщение со сколь угодно малой вероятностью ошибки декодирования, при этом наблюдатель не получит из наблюдаемых данных практически никакой информации о переданном сообщении. В [2] исследована частная модель из [1], когда Σ_m — канал без помех, а Σ_w — канал, по которому наблюдатель в слове (векторе) длины n может просматривать значения координат с номерами из множества $\tau := \{i_1; ...; i_\mu\}$, $(1 \leq \mu \leq n)$. Предполагается, что при фиксированном μ наблюдатель на свое усмотрение определяет состав множества τ, $|\tau| = \mu$. Модель перехвата из работы [2] описывает, например, утечку данных с одного сервера в системе распределения секрета среди нескольких серверов. С целью повышения неопределённости на стороне наблюдателя при перехвате μ произвольных символов в слове длины n в [2] применён метод случайного кодирования, основанный на использовании некоторого линейного кода C и не требующий применения секретных ключей. Отметим, что этот метод впервые рассмотрен в [1] и позже в отечественной публикации был назван методом кодового зашумления [3].

В настоящей работе рассматривается модель случайного перехвата в канале без помех, когда наблюдателю заранее не известно, какие координаты передаваемого вектора он перехватит. В рамках такой модели наблюдатель не может выбирать множество

[1] *Южный федеральный университет, Ростов-на-Дону, Россия*
Контактный e-mail: itaim@mail.ru

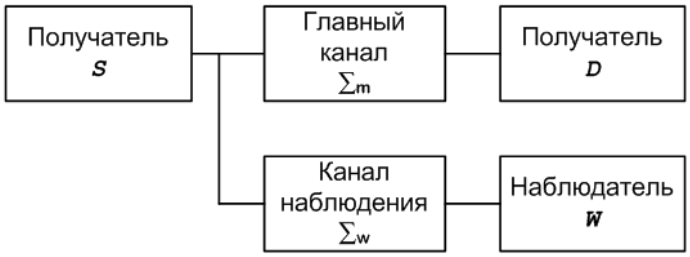

Рис. 1. Схема канала с наблюдением

наблюдаемых координат τ мощности μ, однако ему заранее известна вероятность перехвата для каждого из наборов τ. Эта модель, например, подходит для описания утечки данных с одного сервера в распределённых хранилищах данных [4], где перед записью на сервер порция данных из записываемого в хранилище блока выбирается случайно с некоторой вероятностью. В общем случае распределение вероятностей на конечном множестве наборов перехватываемых координат может быть произвольным, поэтому в качестве меры стойкости кодового зашумления рассматривается энтропия ансамбля информационных сообщений при условии частичной известности кодовых сообщений. Отметим, что в литературе, как правило, решается задача асимптотической оценки меры стойкости в канале со случайным перехватом, в представленной работе, напротив, решается задача неасимптотической оценки. Для произвольного вероятностного распределения и произвольного линейного кода получена формула вычисления меры стойкости, а для равновероятного случайного перехвата μ координат и перехвата, моделируемого каналом со стираниями, показано, что мера стойкости может быть вычислена с помощью введённой в работе характеристики линейного кода.

1. Модель случайного перехвата

1.1. Метод кодового зашумления

Пусть выходом источника $\mathcal{S}$ являются элементы поля Галуа $\mathbb{F} = \mathbb{F}_q$, $q = p^r$, где p — простое число, $r \in \mathbb{N}$. Если генерируемые источником $\mathcal{S}$ данные перед поступлением в канал не преобразуются, то наблюдатель по подслушанным μ элементам переданного слова длины n может составить множество претендентов N_τ путём перебора всех значений неизвестных координат, номера которых принадлежат множеству $\{1;...;n\}\backslash\tau$, где τ — множество номеров наблюдаемых координат, $|\tau| = \mu$. Поскольку, по условию, источник $\mathcal{S}$ равновероятный, то все $q^{n-\mu}$ кандидатов в N_τ равновероятны. Таким образом, если перед поступлением в канал данные источника не преобразуются, то наблюдатель будет иметь ненулевую информацию о переданном слове, так как мощность множества претендентов N_τ строго меньше мощности множества всех сообщений длины n над полем Галуа $\mathbb{F}$. Опишем метод случайного кодирования (кодового зашумления) из [2], который даёт возможность увеличить неопределённость наблюдателя при перехвате μ координат.

Множество натуральных чисел от 1 до n обозначим $\underline{n}$. Для $\tau = \{t_1;...;t_\mu\} \subseteq \underline{n}$ вектор вида $(x_{t_1},...,x_{t_\mu})$, составленный из значений координат вектора $\mathbf{x} = (x_1,...,x_n)$, обозначим $\mathbf{x}_\tau$. Пусть G — порождающая $((n-k)\times n)$-матрица кода C, G^* — $(k\times n)$-матрица ранга k, составленная из векторов множества $\mathbb{F}^n\backslash C$, такая, что $(n\times n)$-матрица вида

$$\widetilde{G} = \begin{pmatrix} G^* \\ G \end{pmatrix}$$

имеет максимальный ранг $\mathrm{rank}(\widetilde{G}) = n$. Тогда кодирование информационного блока $\mathbf{s}(\in \mathbb{F}^k)$ выполняется по правилу $\mathbf{x} = (\mathbf{s}, \mathbf{v})\widetilde{G} = \mathbf{s}G^* + \mathbf{v}G$, где $\mathbf{v} \in \mathbb{F}^{n-k}$ — случайно выбранный вектор. Таким образом, одному информационному блоку длины k соответствует подмножество кодовых блоков мощности q^{n-k}, при этом любым двум разным информационным блокам соответствуют непересекающиеся подмножества кодовых блоков. Нетрудно проверить, что для снятия кодового зашумления с вектора $\mathbf{x}$ можно выбрать такую проверочную $(k \times n)$-матрицу H кода C, для которой выполняется равенство $\mathbf{s} = \mathbf{x}H^T$. Код C, как и в [5], назовём базовым, а код с порождающей матрицей $\widetilde{G}$ — факторным. Парой $(\mathbb{F}^n, C)$ будем обозначать факторный код, построенный по коду C. Отметим, что особенностями метода кодового зашумления являются высокая скорость операций прямого и обратного преобразования и отсутствие необходимости в использовании секретных ключей.

1.2. Общая модель случайного перехвата

Пусть $\mathcal{S}$ — случайный и равновероятный источник, k — длина информационного блока, $(\mathbb{F}^n, C)$ — факторный код, построенный по базовому $(n, n-k)$-коду C, а по каналу Σ_m передаются кодовые векторы факторного кода. Пусть $\mathcal{T}$ — множество всех подмножеств множества $\underline{n}$, $p(\tau)$ — вероятностное распределение, заданное на множестве $\mathcal{T}$. Будем считать, что ансамбль $\{\mathcal{T}, p(\tau))\}$ задает модель случайного перехвата. Другими словами, наблюдатель может просматривать координаты с номерами из множества τ с вероятностью $p(\tau)$. Символом $\mathbf{T}$ обозначим случайный вектор, принимающий значения из $\mathcal{T}$ в соответствии с распределением $p(\tau)$. Ненаблюдаемые координаты назовём стёртыми и вместо них будем писать символ "$*$". Введём обозначения: $\widetilde{\mathbb{F}} := \mathbb{F} \bigcup \{*\}$, $\widetilde{\mathbb{F}}^n := \{\mathbf{x} = (x_1, ..., x_n) : x_i \in \widetilde{\mathbb{F}}\}$, $\mathrm{supp}(\mathbf{x}) := \{i | x_i \neq *\}$, $w(\mathbf{x}) := |\mathrm{supp}(\mathbf{x})|$, $\widetilde{\mathbb{F}}^n(\tau) = \{\mathbf{x} \in \widetilde{\mathbb{F}}^n : \mathbf{x}_\tau \in \mathbb{F}^{|\tau|}, w(\mathbf{x}) = |\tau|\}$.

Как и в работе [2], назовём векторы $\mathbf{x} \in \mathbb{F}^n$ и $\mathbf{z} \in \widetilde{\mathbb{F}}^n$ совместимыми, если $\mathbf{x}_{\mathrm{supp}(\mathbf{z})} = \mathbf{z}_{\mathrm{supp}(\mathbf{z})}$. Совместимость двух векторов $\mathbf{x}$ и $\mathbf{z}$ будем обозначать $\mathbf{x} \sim \mathbf{z}$. Пусть

$$\delta_{\mathbf{x},\mathbf{z}} = \begin{cases} 1, & \mathbf{x} \sim \mathbf{z}, \\ 0, & \text{иначе.} \end{cases}$$

Тогда очевидно, что

$$p(\mathbf{z}|\mathbf{x}) = \delta_{\mathbf{x},\mathbf{z}} \mathrm{Pr}(\mathbf{T} = \mathrm{supp}(\mathbf{z})).$$

Следовательно,

$$p(\mathbf{z}) = \sum_{\mathbf{x} \in \mathbb{F}^n} p(\mathbf{z}, \mathbf{x}) = \sum_{\mathbf{x} \in \mathbb{F}^n} p(\mathbf{z}|\mathbf{x}) p(\mathbf{x}) = \mathrm{Pr}(\mathbf{T} = \mathrm{supp}(\mathbf{z})) \sum_{\mathbf{x} \in \mathbb{F}^n} \delta_{\mathbf{x},\mathbf{z}} p(\mathbf{x}). \tag{1}$$

Случайный вектор, принимающий значения из $\widetilde{\mathbb{F}}^n$ в соответствии с распределением (1), обозначим $\mathbf{Z}$. Так как $p(\mathbf{s}) = 1/q^k$ для всех $\mathbf{s} \in \mathbb{F}^k$ и случайный аргумент при факторном кодировании выбирается случайно и равновероятно, то $p(\mathbf{x}) = 1/q^n$ для всех $\mathbf{x} \in \mathbb{F}^n$. В этом случае (1) примет вид

$$p(\mathbf{z}) = \frac{\mathrm{Pr}(\mathbf{T} = \mathrm{supp}(\mathbf{z}))}{q^{w(\mathbf{z})}}. \tag{2}$$

Пусть $\mathbf{S}$ — случайный вектор, принимающий случайно и равновероятно значения из $\mathbb{F}^k$ и моделирующий информационные блоки длины k, сгенерированные источником $\mathcal{S}$. Мерой стойкости факторного кода при случайном частичном перехвате будем считать величину

$$\Delta^{\mathrm{rand}}(C) = H(\mathbf{S}|\mathbf{Z}), \tag{3}$$

где $\mathbf{Z}$ принимает значения из $\widetilde{\mathbb{F}}^n$ в соответствии с (2), а $H(\mathrm{A}|\mathrm{B})$ — условная энтропия случайной величины A при условии наличия информации о случайной величине B. При $\Delta = k$ полагают, что обеспечивается совершенная защита, или полная неопределённость наблюдателя.

1.3. Теорема о вычислении меры стойкости $\Delta^{\mathrm{rand}}(C)$

Для полноты изложения кратко приведём результат оценки из работы [2] меры стойкости (3), когда наблюдатель на своё усмотрение выбирает μ координат для подслушивания. Если $\mathbf{c}$ — кодовый блок, соответствующий информационному блоку $\mathbf{s}$ длины k и передаваемый по каналу Σ_m, а $\mathbf{z}$ — блок длины n, у которого значения координат с номерами из множества τ ($|\tau| = \mu$) совпадают со значениями соответствующих координат вектора $\mathbf{c}$, а значения координат из множества $\{1; ...; n\}\backslash\tau$ неизвестны, то через $N_\tau(\mathbf{z})$ обозначим множество информационных блоков, которое соответствует множеству кодовых блоков мощности $q^{n-\mu}$, полученному перебором неизвестных координат в блоке $\mathbf{z}$. Очевидно, что $|N_\tau(\mathbf{z})| \leq q^k$ и $\mathbf{s} \in N_\tau(\mathbf{z})$. В [2] показано, что мера стойкости (3) в случае возможности выбора наблюдаемых координат вычисляется как логарифм (по основанию q) минимума мощности списка $N_\tau(\mathbf{z})$ по всем $\mathbf{z}$ и всем τ мощности μ. Меру стойкости для модели из [2] обозначим $\Delta(C, \mu)$. В работе [6] для модели наблюдения из [2] найдено однозначное соответствие между весовой иерархией кода C и $\Delta(C, \mu)$. В [7] показано, что для оценки $\Delta(C, \mu)$ также удобно использовать обратный профиль отношений размерности к длине (Inverse Dimension/Length Profile — IDLP) кода, дуального к базовому. Далее будем использовать обобщение понятия IDLP, поэтому приведём необходимые определения из [7]. В соответствии с [7], IDLP — это последовательность чисел

$$\widetilde{k}_i(C) = \min_{\tau \subseteq \underline{n}: |\tau| = i} \{dim(P_\tau[C])\}, \quad i = 0, ..., n,$$

где $P_\tau[C] = \{\mathbf{c}_\tau : \mathbf{c} \in C\}$. В работе [7], в частности, показано, что для модели из [2]

$$\Delta(C, \mu) = \widetilde{k}_{n-\mu}(C^\perp),$$

где $C^\perp$ — код, дуальный к базовому коду C.

Следующая теорема обобщает результат [2] на случай произвольной стратегии частичного перехвата.

Теорема. *Пусть $(\mathbb{F}^n, C)$ — факторный код, построенный по базовому $(n, n-k)$-коду C. Тогда*

$$\Delta^{\mathrm{rand}}(C) = \mathrm{E}_{\mathbf{T}}[\mathrm{rank}(H_{\overline{\mathbf{T}}})], \tag{4}$$

где $\mathrm{E}_\phi(\xi(\phi))$ — математическое ожидание случайной величины $\xi(\phi)$, зависящей от случайной величины ϕ, $\overline{\mathbf{T}} = \underline{n}\backslash\mathbf{T}$, $H_{\overline{\mathbf{T}}}$ — проекция матрицы H на множество $\overline{\mathbf{T}}$, состоящая из столбцов матрицы H с номерами из $\overline{\mathbf{T}}$.

Доказательство.

$$\Delta^{\text{rand}}(C) = H(\mathbf{S}|\mathbf{Z}) = -\sum_{\mathbf{z}\in\widetilde{\mathbb{F}}^n}\sum_{\mathbf{s}\in\mathbb{F}^k} p(\mathbf{s},\mathbf{z})\log p(\mathbf{s}|\mathbf{z}) = -\sum_{\mathbf{z}\in\widetilde{\mathbb{F}}^n}\sum_{\mathbf{s}\in\mathbb{F}^k} p(\mathbf{z})p(\mathbf{s}|\mathbf{z})\log p(\mathbf{s}|\mathbf{z}) =$$

$$= -\sum_{\mathbf{z}\in\widetilde{\mathbb{F}}^n}\sum_{\mathbf{s}\in\mathbb{F}^k} \frac{\Pr(\mathbf{T}=\text{supp}(\mathbf{z}))}{q^{w(\mathbf{z})}} p(\mathbf{s}|\mathbf{z})\log p(\mathbf{s}|\mathbf{z}).$$

Так как для любого $\mathbf{z}\in\widetilde{\mathbb{F}}^n$ выполняется равенство $p(\mathbf{z}|\text{supp}(\mathbf{z})) = 1/q^{w(\mathbf{z})}$, то

$$\Delta^{\text{rand}}(C) = -\sum_{\tau\in\mathcal{T}}\sum_{\mathbf{y}\in\widetilde{\mathbb{F}}^n(\tau)}\sum_{\mathbf{s}\in\mathbb{F}^k} \Pr(\mathbf{T}=\tau)p(\mathbf{y})p(\mathbf{s}|\mathbf{y})\log p(\mathbf{s}|\mathbf{y}) =$$

$$= -\sum_{\tau\in\mathcal{T}} \Pr(\mathbf{T}=\tau)\sum_{\mathbf{y}\in\widetilde{\mathbb{F}}^n(\tau)}\sum_{\mathbf{s}\in\mathbb{F}^k} p(\mathbf{y})p(\mathbf{s}|\mathbf{y})\log p(\mathbf{s}|\mathbf{y}).$$

Заметим, что для фиксированного τ величина $-\sum\limits_{\mathbf{y}\in\widetilde{\mathbb{F}}^n(\tau)}\sum\limits_{\mathbf{s}\in\mathbb{F}^k} p(\mathbf{y})p(\mathbf{s}|\mathbf{y})\log p(\mathbf{s}|\mathbf{y})$ является энтропией наблюдателя при перехвате координат из множества τ. Согласно [2], эта энтропия равна $\text{rank}(H_{\overline{\tau}})$. Таким образом,

$$\Delta^{\text{rand}}(C) = \sum_{\tau\in\mathcal{T}} \Pr(\mathbf{T}=\tau)\text{rank}(H_{\overline{\tau}}),$$

откуда следует (4). □

1.4. Модель случайного и равновероятного перехвата μ координат

Рассмотрим модель $\{\mathcal{T}, p(\tau))\}$ случайного и равновероятного перехвата μ координат, где

$$p(\tau) = \begin{cases} \dfrac{1}{C_n^\mu}, & \text{если } |\tau| = \mu, \\ 0, & \text{иначе.} \end{cases}$$

Для (n,k)-кода C и числа $i(\in \underline{n})$ набор из $k+1$ чисел вида

$$(u_{i,0}(C), ..., u_{i,k}(C)), \tag{5}$$

где $u_{i,j} = |\{P_\tau[C] : |\tau| = i, \dim(P_\tau[C]) = j\}|$, назовем спектром IDLP. Пусть $\Delta^{\text{unif}}(C,\mu) = H(\mathbf{S}|\mathbf{Z})$ — неопределённость наблюдателя при равновероятном перехвате μ координат. Тогда из (4) и (5) имеем нижеприведённое следствие.

Следствие 1. *Пусть $(\mathbb{F}^n, C)$ — факторный код, построенный по базовому $(n, n-k)$-коду C, а наблюдатель подслушивает μ координат случайно и равновероятно. Тогда*

$$\Delta^{\text{unif}}(C,\mu) = \frac{1}{C_n^\mu}\sum_{i=0}^{k} i \times u_{n-\mu,i}(C^\perp). \tag{6}$$

Доказательство. Из (4) и определения спектра IDLP получим

$$\Delta^{\text{unif}}(C,\mu) = H(\mathbf{S}|\mathbf{Z}) = \sum_{\tau\in\mathcal{T}} p(\tau)\text{rank}(H_{\overline{\tau}}) =$$

$$= \frac{1}{C_n^\mu}\sum_{\tau\in\mathcal{T}} \text{rank}(H_{\overline{\tau}}) = \frac{1}{C_n^\mu}\sum_{i=0}^{k} i \times u_{n-\mu,i}(C^\perp).$$

□

1.5. Модель перехвата по симметричному каналу со стираниями EC(p)

Рассмотрим модель перехвата по симметричному каналу со стираниями EC(p), где p — вероятность стирания одного символа. В соответствии с введённым выше определением перехват по такому каналу моделируется ансамблем $\{\mathcal{T}, p(\tau))\}$, где $p(\tau) = (1-p)^{|\tau|}p^{n-|\tau|}$. Неопределённость наблюдателя, перехватывающего кодовые слова факторного кода $(\mathbb{F}^n, C)$ по каналу со стираниями EC(p), обозначим $\Delta^{\mathrm{EC}(p)}(C)$. По определению, $\Delta^{\mathrm{EC}(p)}(C) = H(\mathbf{S}|\mathbf{Z})$, где случайный вектор $\mathbf{Z}$ в соответствии с (2) имеет распределение вида

$$p(\mathbf{z}) = \frac{(1-p)^{w(\mathbf{z})}p^{n-w(\mathbf{z})}}{q^{w(\mathbf{z})}}.$$

Следствие 2. *Пусть канал наблюдателя моделируется симметричным каналом со стираниями* EC(p) *с вероятностью* p, M *— случайная величина, принимающая значения из множества* $\{0; 1; ...; n\}$ *с вероятностью* $\Pr(M = \mu) = C_n^\mu p^{n-\mu}(1-p)^\mu$. *Тогда*

$$\Delta^{\mathrm{EC}(p)}(C) = \mathrm{E}_M\left[\Delta^{\mathrm{unif}}(C, M)\right]. \tag{7}$$

Доказательство. Из (4) имеем

$$\Delta^{\mathrm{EC}(p)}(C) = \sum_{\tau\in\mathcal{T}} \Pr(\mathbf{T}=\tau)\mathrm{rank}(H_{\overline{\tau}}) = \sum_{\mu=0}^{n}\sum_{\tau:|\tau|=\mu} \Pr(\mathbf{T}=\tau)\mathrm{rank}(H_{\overline{\tau}}) =$$

$$= \sum_{\mu=0}^{n}\sum_{\tau:|\tau|=\mu} (1-p)^\mu p^{n-\mu}\mathrm{rank}(H_{\overline{\tau}}) = \sum_{\mu=0}^{n}(1-p)^\mu p^{n-\mu}\sum_{i=0}^{k} u_{n-\mu,i}(C)^\perp =$$

$$= \sum_{\mu=0}^{n} C_n^\mu (1-p)^\mu p^{n-\mu}\frac{1}{C_n^\mu}\sum_{i=0}^{k} u_{n-\mu,i}(C)^\perp = \mathrm{E}_M\left[\Delta^{\mathrm{unif}}(C, M)\right].$$

□

2. Примеры вычисления меры стойкости при случайном перехвате

Для исследования меры стойкости различных кодов разработан и реализован алгоритм подсчёта спектра IDLP. С помощью этой программы, в частности, посчитаны спектры IDLP для кода, дуального коду Хэмминга, кода, дуального коду Рида — Маллера и некоторых кодов, дуальных к низкоплотностным кодам. На основе вычисленных спектров IDLP по формуле (6) из следствия 1 и формуле (7) из следствия 2 вычислены соответствующие меры стойкости. Для примера результат вычисления спектра IDLP для кода, дуального коду Рида — Маллера $\mathcal{RM}(1,3)$, приведён в таблице. На рис. 2 для каждого из четырёх рассмотренных кодов построены по три графика: зависимости $\Delta(C,\mu)$ и $\Delta^{\mathrm{unif}}(C,\mu)$ от количества перехватываемых координат μ (штриховая и сплошная линии) и асимптотическая зависимость неопределённости при заданных соотношениях параметров n, k и μ (пунктирная линия). На рис. 2, *д* приведены нормированные графики зависимостей $\Delta^{\mathrm{unif}}(C,\mu)/k$ от доли перехвата для всех четырёх кодов: штрихпунктирной линией показан код, дуальный коду Рида — Маллера $\mathcal{R}(1,3)$, пунктирной — дуальный коду Хэмминга, сплошной — дуальный LDPC$(2,4)$-коду, штриховой — дуальный LDPC$(2,5)$-коду. Зависимость меры стойкости для канала перехвата, моделируемого двоичным каналом со стираниями, представлена на рис. 3.

Результаты вычисления $\{u_{i,j}(C^{\perp})\}_{i=0,\ldots,n;j=0,\ldots,k}$, $\Delta(C,\mu)$ и $\Delta^{\mathrm{unif}}(C,\mu)$, C — $(8,4)$-код, дуальный коду Рида — Маллера $\mathcal{R}(1,3)$

Параметры		j					$\Delta(C,\mu)$	$\Delta^{\mathrm{unif}}(C,\mu)$
		0	1	2	3	4		
i, $8-\mu$	0	1	0	0	0	0	0	0
	1	0	8	0	0	0	1	1
	2	0	0	28	0	0	2	2
	3	0	0	0	56	0	3	3
	4	0	0	0	14	56	3	3.8
	5	0	0	0	0	56	4	4
	6	0	0	0	0	28	4	4
	7	0	0	0	0	8	4	4
	8	0	0	0	0	1	4	4

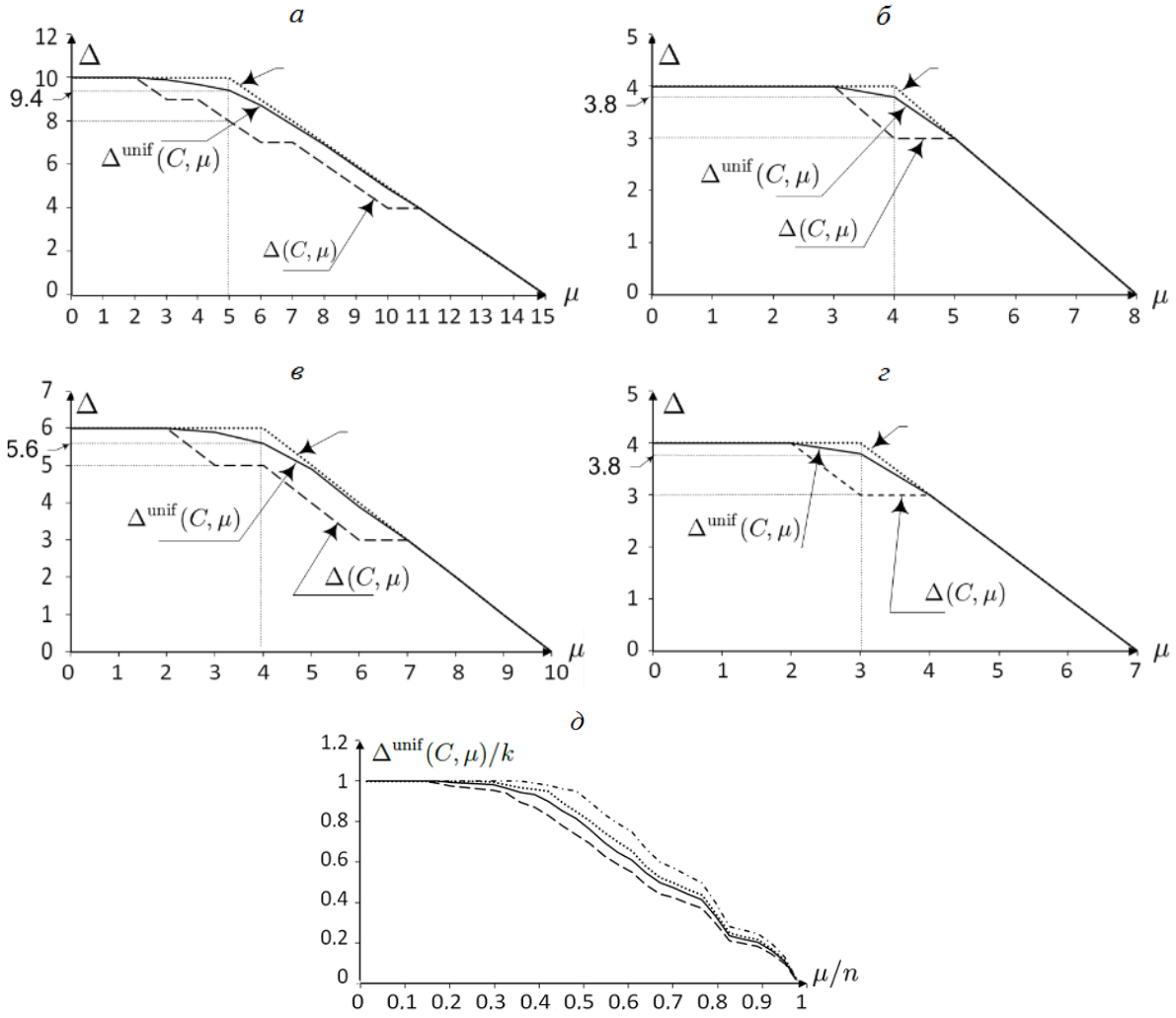

Рис. 2. Зависимости $\Delta(C,\mu)$ (штриховая линия), $\Delta^{\mathrm{unif}}(C,\mu)$ (сплошная линия) и асимптотической меры стойкости (пунктирная линия) от μ: C — код, дуальный LDPC$(2,5)$-коду *(а)*, дуальный коду Рида — Маллера $\mathcal{R}(1,3)$ *(б)*, дуальный LDPC$(2,4)$-коду *(в)*, дуальный коду Хэмминга *(г)* (стрелками сверху указана теоретическая граница при наблюдении μ координат); *д* — сравнение нормированных величин $\Delta^{\mathrm{unif}}(C,\mu)/k$ в зависимости от доли перехвата

2.1. Анализ результатов вычислений

Из приведённого рис. 2, *а – г* видно, что график $\Delta^{\mathrm{unif}}(C,\mu)$ либо совпадает с графиком $\Delta(C,\mu)$, либо выше него. Этот результат объясняется тем, что в модели [2] наблюдатель может заранее подобрать множество номеров перехватываемых координат мощности μ так, чтобы его неопределённость была минимально возможной при заданном μ. В модели равновероятного перехвата у наблюдателя такой возможности выбора нет, поэтому при перехвате при заданном μ в одном случае множество τ может быть таким, что неопределённость будет минимально возможной, в другом — максимально возможной. В силу этого среднее значение неопределённости (меры стойкости) $\Delta^{\mathrm{unif}}(C,\mu)$ при равновероятном перехвате может либо совпадать с величиной $\Delta(C,\mu)$, либо превышать её. На сколько $\Delta^{\mathrm{unif}}(C,\mu)$ больше $\Delta(C,\mu)$ при заданном μ зависит от того, на каком числе наборов τ мощности μ (из всего множества наборов координат мощности μ) ранг матрицы $H_{\bar{\tau}}$ равен $\Delta(C,\mu)$, т. е. зависит от спектра IDLP. В частности, из приведённой таблицы видно, что при перехвате черырёх координат ($\mu = 4$) из 70 наборов имеются только 14 наборов перехватываемых координат, на которых ранг соответствующих матриц равен 3, на остальных же 56 наборах ранг равен 4. Это объясняет высокую неопределённость $\Delta^{\mathrm{unif}}(C,4) = 3.8$ при случайном и равновероятном перехвате против $\Delta(C,4) = 3$ при перехвате с возможностью выбора перехватываемых координат.

Для других кодов (см. рис. 2) высокая мера стойкости, близкая к асимптотической, обеспечивается также при бóльшем уровне перехвата, нежели чем в модели из [2]. Например, из графика на рис. 2, *а* видно, что в модели равновероятного перехвата неопределённость $\Delta^{\mathrm{unif}}(C,\mu)$ близка к максимальному значению при перехвате до четырёх координат, в то же время в модели с выбором наблюдаемых координат [2] максимальное значение $\Delta(C,\mu)$ обеспечивается только при перехвате не более двух координат. С прикладной точки зрения это означает, что если факторный код $(\mathbb{F}^n, C)$, где $C = \mathrm{LDPC}^{\perp}(2,5)$, не позволяет противодействовать наблюдателю, имеющему возможность на свой выбор подслушивать четыре координаты из слова длины 15, то этот же факторный код может быть использован, если наблюдатель случайно может подслушать не более четырёх координат передаваемого вектора длины 15.

Для примера рассмотрим применение факторного кода $(\mathbb{F}^n, C)$, где $C = \mathrm{LDPC}^{\perp}(2,5)$, в рамках модели распределения данных среди N серверов. Пусть перед записью на серверы каждое кодовое слово разбивается фиксированным образом на N частей, где i-часть — значения кодовых координат из множества $W_i (\subseteq \{1;...;15\})$, $i = 1,...,N$, $W_i \cap W_j = \emptyset$ для $i \neq j$. Тогда i-й сервер всегда будет иметь в распоряжении (наблюдать) координаты из множества W_i, $i = 1,...,N$. Данная модель аналогична модели из работы [2], так как заранее известно, какая порция данных будет записана на каждый сервер. Тогда, как следует из рис. 2, *а*, для обеспечения совершенной защиты необходимо, чтобы $|W_i| \leq 2$, $i = 1,...,N$. Поэтому для организации распределённого хранения потребуется не менее восьми серверов. Если же для каждого кодового слова разбиение на равные части выполняется случайным и равновероятным образом, то заранее не известно, какая порция данных кодового слова будет записана на сервер, поэтому данная ситуация может быть описана моделью случайного и равновероятного перехвата. Как следует из рис. 2, *а*, в последнем случае в среднем будет обеспечиваться защита, близкая к совершенной, уже при $|W_i| \leq 4$, $i = 1,...,N$, в силу чего $N \geq 4$, т. е. при таком способе распределения информации уменьшается количество серверов, необходимых для организации распределённого хранилища при сохранении высокой защищённости.

На рис. 2, *д* показано сравнение нормированных величин $\Delta^{\mathrm{unif}}(C,\mu)/k$ в зависимости от доли перехвата. Как следует из приведённых данных, лучшую защитную характеристику имеет код Рида — Маллера $\mathcal{RM}(1,3)$. Этот же код обладает лучшей защитной характеристикой и для случая, когда канал перехвата моделируется двоичным каналом со стираниями (рис. 3). Вместе с тем это не означает, что $\mathcal{RM}(1,3)$ лучше других рассмотренных кодов, так как все они имеют различную скорость передачи информации. Более того, с точки зрения близости неасимтотической оценки, вычисленной в соответствии с (7), к асимптотической неопределённости наилучшим является код LDPC$(2,5)$. Действительно, согласно [1], если скорость передачи не превышает p, где p — вероятность стирания двоичного символа, то найдется такой базовый код, возможно с очень большими значениями параметров k и n, для которого соответствующий факторный код имеет скорость передачи информации, близкую к p, и при этом будет обеспечиваться полная неопределённость у перехватчика: $\Delta^{\mathrm{BEC}(p)}(C)/k \approx 1$. В работе [8], в частности, показано, что при больших k и n этому условию удовлетворяют некоторые слабоплотные коды. На рис. 3 для каждого из четырёх рассмотренных кодов для точек $p = k/n$ (k и n — соответствующие параметры этих кодов) показана реальная (неасимтотическая) неопределённость, вычисленная по формуле (7). Видно, что ближе всех к теоретически возможной неопределённости является код, дуальный LDPC$(2,5)$-коду: его нормированная неопределённость равна 0.904. Поэтому с прикладной точки зрения корректным представляется сравнение (используя полученные в работе формулы) факторных кодов только в случае, когда над полем $\mathbb{F}$ они имеют одинаковые параметры: размерность k и длину n.

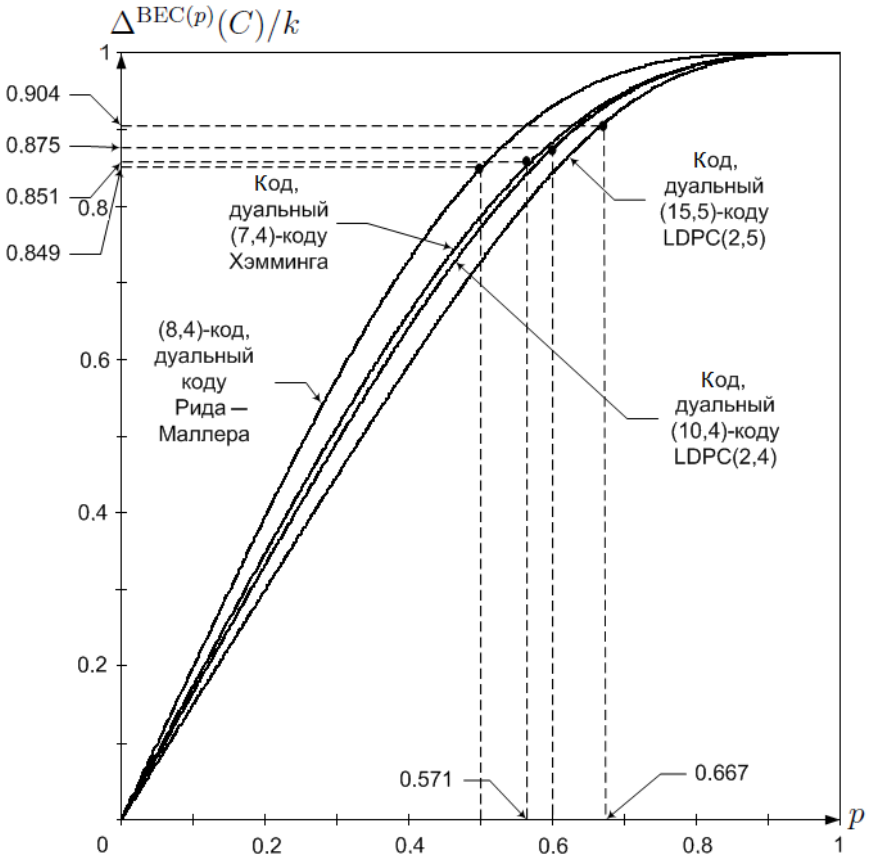

Рис. 3. Вычисление неопределённости для кодового зашумления с каналом перехвата BEC(p)

2.2. Замечание о совершенной защите

Случай, когда обеспечивается полная неопределённость наблюдателя, с точки зрения защиты информации является наиболее подходящим. Но, как следует, например, из рис. 2, такая степень стойкости обеспечивается при малом уровне перехвата. Если же перехватывается число символов, большее того, при котором обеспечивается совершенная защита, то даже при случайном и равновероятном перехвате неопределённость становится строго меньше асимптотически достижимой максимальной неопределённости, равной k. Используя терминологию теории информации, можно сказать, что у наблюдателя в этом случае имеется ненулевая информация о сообщении. С прикладной точки зрения о степени близости неасимптотической неопределённости к асимптотически достижимой и равной k можно судить в том случае, если наблюдатель может извлечь практическую пользу из имеющейся у него ненулевой информации о сообщении.

Например, в работе [9] показано, что если наблюдатель имеет ненулевую информацию о сообщении, то он может попытаться провести атаку повторного перехвата данных для полного снятия неопределённости. Тогда, используя формулу (1.23) из работы [10] и результаты вычислений из вышеприведённой таблицы, получим, что при случайном и равновероятном перехвате четырёх координат для полного снятия неопределённости потребуется 20 повторных перехватов. Заметим, что для полного снятия неопределённости при перехвате четырёх координат на выбор наблюдателя достаточно четырёх повторных перехватов кодовых слов [9]. Аналогично для кода, дуального коду $\mathrm{LDPC}(2,5)$, используя ту же формулу из [10], получим, что для полного снятия неопределённости при случайном и равновероятном перехвате трёх координат потребуется 101 повторный перехват, а при перехвате четырёх координат необходимо 35 повторных перехватов. Тот же анализ можно провести и для канала перехвата, моделируемого каналом $\mathrm{BEC}(p)$. Например, из рис. 3 видно: при $p = 0.667$ нормированная неопределённость для $\mathrm{LDPC}^{\perp}(2,5)$ равна 0.904. Это означает, что для полного снятия неопределённости потребуется порядка 12 повторных перехватов, причём асимптотически при $p = 0.667$ обеспечивается совершенная защита и любое количество повторных перехватов не повлияет на уже достигнутый результат. Следует отметить, что конкретные количества перехватов для полного снятия неопределённости можно оценить только в том случае, когда известен контекст прикладной задачи: тип канала передачи, объём передаваемых данных, наличие повторяющихся информационных блоков в сообщении и т. п. В общем случае в рамках модели многократного перехвата, если при заданных уровне перехвата или вероятности стирания символа для полного снятия неопределённости требуется N повторных перехватов при $N >> N_w$, где N_w — максимальное количество повторных перехватов, которые может осуществить наблюдатель, то даже при необеспечиваемой совершенной стойкости в теоретико-информационном смысле можно говорить о высокой практической стойкости кодового зашумления. Величина N_w в этой ситуации существенно зависит от контекста прикладной задачи.

Таким образом, прикладное значение защитных характеристик определяется контекстом прикладной задачи и тем, как наблюдатель может воспользоваться имеющейся у него информацией о закодированном сообщении.

Заключение

В работе получена формула вычисления неасимптотической меры стойкости кодового зашумления для модели случайного перехвата. Для частных случаев, когда случайный

перехват описывается моделями равновероятного случайного перехвата и перехвата, моделируемого симметричным каналом со стираниями, получены формулы для вычисления меры стойкости с использованием спектра IDLP кода, дуального базовому коду. С прикладной точки зрения представленные формулы позволяют при использовании разработанной в [10] теории оценить практическую стойкость кодового зашумления к атаке повторного перехвата и к коалиционной атаке, что актуально, например, в распределённых хранилищах данных, построенных на базе множества независимых серверов (носителей данных).

Результаты вычислений для параметров $k = 4$, $n = 8$ и $\mathbb{F} = \mathbb{F}_2$ показали, что для канала со случайным перехватом и для канала перехвата, моделируемого двоичным каналом со стираниями, наилучшие защитные характеристики имеет базовый код, дуальный коду Рида — Маллера $\mathcal{RM}(1, 3)$. Учитывая, что коды Рида — Маллера применяются в задачах защиты информации достаточно широко (см. [4, 11, 12]), отмеченный факт подчёркивает высокое прикладное значение кодов Рида — Маллера в задачах защиты информации.

Отметим, что теоретический и практический интерес представляет нахождение формульных выражений для вычисления спектра IDLP для конкретных кодов, так как переборное решение данной задачи существенно усложняется с ростом k и n.

Список литературы

[1] Wyner A.D. The wire-tap channel // Bell System Techn. J. 1975. Vol. 54, No. 8. P. 1355–1387.

[2] Ozarov L.H., Wyner A.D. Wire-tap channel II // Ibid. 1984. Vol. 63. P. 2135–2157.

[3] Коржик В.И., Яковлев В.А. Неасимптотические оценки эффективности кодового зашумления одного канала // Проблемы передачи информации. 1981. Т. 17, вып. 4. С. 11–18.

[4] Косолапов Ю.В., Никулин В.Э. Способ организации распределённого хранилища, устойчивого к частичной утечке данных // Материалы XIII Междунар. научно-практ. конф. “ИБ-2013”. Ч. I. Таганрог: Изд-во ЮФУ, 2013. С. 186–191.

[5] Деундяк В.М., Косолапов Ю.В. Математическая модель канала с перехватом второго типа // Изв. высших учебных заведений. Северо-Кавказский регион. Естественные науки. 2008. № 3(145). С. 3–8.

[6] Wei V.K. Generalized hamming weights for linear codes // IEEE Trans. on Inform. Theory. 1991. Vol. 37, No. 5. P. 1412–1418.

[7] Forney G.D. Dimension/length profiles and trellis complexity of linear block codes // Ibid. 1994. Vol. 40, No. 6. P. 1741–1752.

[8] Rathi V., Andersson M., Thobaben R. et al. Performance Analysis and Design of Two Edge Type LDPC Codes for the BEC Wiretap Channel. arXiv:1009.4610v2 [cs.IT]. Последнее обращение 04.03.2014.

[9] Деундяк В.М., Косолапов Ю.В. Об одном методе снятия неопределённости в канале с помехами в случае применения кодового зашумления // Изв. ЮФУ. Технические науки. 2014. С. 197–208.

[10] Шанкин Г.П. Ценность информации. Вопросы теории и приложений. М.: Филоматис, 2004. 128 с.

[11] Циммерман К.-Х. Методы теории модулярных представлений в алгебраической теории кодирования. М.: МЦНМО, 2011. 246 с.

[12] Sidelnikov V.M. A public-key cryptosystem based on binary Reed — Muller codes // Discrete Math. and Appl. 1994. No. 4(3). P. 191–207.

Поступила в редакцию 28 апреля 2014 г.,
с доработки — 15 августа 2014 г.

Computation of the measure of resistance for Code Noising in a channel with random partial interception

Kosolapov Jury V.[1], Kurchev Nuri O.[1]

Purpose. Generalize partial interception models discussed earlier, and to find the characteristics of linear codes, which determine the measure for resistance of the code noising.
Methodology. We use the traditional methods of information theory and linear codes.
Findings. The generalized model of random wire-tap II channel is constructed for the case, when the error-free main channel is protected by the code noising method. For the constructed model, the formula for the calculation of measures of resistance is obtained. It was found that the measure of resistance in general is the expectation of ranks for the projections of the check matrix of the base code to admissible sets of intercepted coordinates. Further, a more generalized model of interception is considered for two particular cases: a fixed number of equiprobable interception coordinates and the interception, the simulated channel with erasures with a fixed probability of erasure. In these cases, the formula for calculation the measures of resistance is greatly simplified when one of the spectral characteristics of the base code is considered. For the both of these special cases, the results of the calculation of measures of resistance in the case of some well-known codes are given. In particular, the results showed that a perfect security at random intercept of fixed number of coordinates may be achieved even at a greater quantity of observed coordinates compared to the model of initially known interception coordinates discussed by L.H. Ozarov and A.D. Wainer.
Scope results. The results can be used in communications and distributed data storage systems, where the code noising is used to protect the data.
Originality/value. The most interesting result is in the simplification of formulas for the calculation of measures of resistance to random interception when other special cases of interception along with the examination of specific basic codes are considered.
Keywords: a random intercept, code noising, observer, uncertainty.

Received 28 April 2014
Received in revised form 15 August 2014

[1]*South Federal University, Rostov-on-Don, 344006, Russia*
Corresponding author: Kosolapov Jury V., e-mail: itaim@mail.ru