Экспериментальное исследование корректирующей способности матричного метода равновесных столбцов защиты данных от стираний

Е.Е. Айдаркин [1], Н.С. Могилевская [1]
[1] ФГАОУ ВО «Южный федеральный университет»,
344006, Ростовская обл., г. Ростов-на-Дону, ул. Большая Садовая, 105/42

Аннотация

В работе рассматриваются алгебраические способы защиты данных при их передаче по стирающему каналу. Стирания в каналах рассматриваются двух видов: независимые и группирующиеся. Для организации группирующихся стираний модифицирована модель Гилберта генерации потока ошибок. В качестве методов защиты данных от стираний используются метод равновесных столбцов и его модификация, позволяющая в некоторых случаях упростить процесс декодирования. Создано программное средство, реализующее имитационную модель двоичного помехоустойчивого канала с возможностью выбора типа стираний и метода защиты. С помощью этой модели проведено экспериментальное исследование корректирующей способности рассматриваемых методов. Показано, что группирующиеся стирания уменьшают вероятность успешного декодирования для обоих методов и их различных входных параметров по сравнению с независимыми стираниями. Проанализированы преимущества и недостатки метода равновесных столбцов и его модификации. Предложен способ борьбы с группирующимися стираниями за счет использования дополнительной избыточности. Для рассматриваемых методов защиты данных в каналах с независимыми стираниями предложена теоретическая оценка неверного декодирования, основанная на векторе вероятностей успешного декодирования. Предложен способ применения этой оценки для случая группирующихся стираний.

Ключевые слова: стирание, помехоустойчивый канал передачи данных, группирующиеся стирания, модель Гилберта, метод равновесных столбцов.

Цитирование: **Айдаркин, Е.Е.** Экспериментальное исследование корректирующей способности матричного метода равновесных столбцов защиты данных от стираний / Е.Е. Айдаркин, Н.С. Могилевская // Компьютерная оптика. – 2022. – Т. 46, № 5. – С. 840-847. – DOI: 10.18287/2412-6179-CO-1122.

Citation: Aydarkin EE, Mogilevskaya NS. Experimental study of a matrix method of equal-weight columns correcting ability to protect data from erasure. Computer Optics 2022; 46(5): 840-847. DOI: 10.18287/2412-6179-CO-1122.

Введение

Передаваемые по каналам связи данные могут быть повреждены ошибками. Использование в системах связи помехоустойчивых методов является стандартной практикой для защиты данных от искажений [1 – 5]. В этой работе рассматриваются каналы со стираниями. Стирание – это разновидность ошибки, при которой получатель принимает слово с верными символами и с номерами координат, в которых произошла потеря символа [6]. Значения таких координат обычно помечаются символом *. Такой тип ошибок можно встретить, например, в распределенных хранилищах, RAID-массивах, каналах и сетях передачи данных [6]. Для борьбы со стираниями в каналах связи применяются, например, такие методы кодирования, как Аль-Шахи-Илова [7], Пана [8], метод равновесных столбцов [1]; в сетях передачи данных для противодействия стираниям используют, например, ранговые [2, 9] и каскадные коды [10].

Известно, что в случае каналов с ошибками, в которых используются алгебраические методы помехоустойчивого кодирования для защиты данных от искажений, на корректирующую способность кодеков значительное влияние имеет не только интенсивность ошибок, но и их структура [6]. Ошибки могут группироваться в пакеты и более сложные структуры ошибок [11, 12]. Существует большое количество моделей потоков группирующихся ошибок, каждая из которых является адекватной для описания ошибок в каналах определенного типа. Однако при исследовании корректирующей способности методов борьбы со стираниями обычно предполагается, что стирания происходят в каналах независимо и равномерно.

В данной работе рассматриваются два метода защиты данных от стираний: метод равновесных столбцов (метод РС) и его модификация (метод МРС) [1]; проводится исследование корректирующей способности этих методов по отношению к равномерным независимым стираниям и группирующимся стираниям. Для моделирования группирующихся стираний известная модель Гилберта [3] генерации потока группирующихся ошибок модифицирована для случая стираний. Построена имитационная модель сти-

рающего канала передачи, с ее помощью проведены эксперименты по сравнению корректирующей способности методов РС и МРС по отношению к стираниям различной структуры.

1. Модель канала передачи данных с защитой от стираний

Основная идея помехоустойчивого кодирования состоит во внесении специально организованной избыточности в передаваемые данные, что затем позволяет по полученным из канала зашумленным данным восстановить исходные. Будем считать, что передаваемые данные являются элементами поля Галуа F_q мощности q. Параметры кода C обычно описываются тройкой $[n,k,d]_q$, где n – длина кода, k – размерность кода ($n>k$), d – кодовое расстояние; если кодовое расстояние неизвестно, то можно говорить о $[n,k]_q$-коде. Расстояние d является характеристикой качества кода и влияет на количество стираний t, исправляемых в одном кодовом векторе:

$$d \geq t+1. \tag{1}$$

Отношение n/k называется избыточностью кода. Увеличение избыточности, как правило, повышает количество исправляемых стираний. Но с увеличением избыточности снижается скорость кода – отношение $r=k/n$, что ведет к возрастанию времени передачи данных по каналу [6, 13].

Рассмотрим модель передачи данных по каналу, который вносит в передаваемые данные непреднамеренные ошибки типа стирания, а для защиты данных в модели используется помехоустойчивый код C.

Источник сообщений выдает информационные векторы вида $\overline{m}=(m_1,m_2,\ldots,m_k)\in F_q^k$. Далее эти сообщения поступают на вход кодера канала, в котором используется алгебраический помехоустойчивый $[n,k,d]_q$ -код C, исправляющий стирания. На выходе кодера формируются кодовые векторы $\overline{c}=(c_1,c_2,\ldots,c_n)\in F_q^n$. Таким образом, можно задать оператор кодирования

$$C: F_q^k \to F_q^n.$$

Затем кодовые векторы поступают в линию связи, где под влиянием искажений символы кодовых слов могут быть стерты, т.е. заменены на символ *. Действия стираний можно описать аддитивным законом

$$\overline{c}+\overline{e}=(c_1+e_1,c_2+e_2,\ldots,c_n+e_n), \tag{2}$$

где $\overline{e}=(e_1,e_2,\ldots,e_n)$, $e_i\in\{0,*\}$ – вектор стираний, $\overline{c}$ $\left(\in F_q^n\right)$ – кодовый вектор. При этом $\forall a\in F_q: a+*=*+a=*$. Составим оператор линии связи L:

$$L:\ F_q^n \to F_{q^*}^n,\ F_{q^*}=F_q\cup\{*\}.$$

Задачей декодера канала является восстановление информационного сообщения $\overline{m}'\in F_q^k$ из зашумленного кодового вектора $\overline{c}'=\overline{c}+\overline{e}$, полученного из линии связи. Если количество стираний t в $\overline{c}'$ удовлетворяет неравенству (1), то декодер корректно восстанавливает исходное информационное сообщение $\overline{m}$, т.е. $\overline{m}'=\overline{m}$. Иначе, если количество стертых символов $t>d-1$, то процесс восстановления информационного вектора декодером завершается сообщением об ошибке, которое может быть представлено вектором E_d=(*,*,…,*) длины k. Тогда оператор декодера может быть задан следующим образом:

$$D: F_{q^*}^n \to F_q^k \cup\{E_d\}.$$

Совокупность описанных операторов определяет канал с защитой от стираний:

$$D\cdot L\cdot C: F_q^k \to F_q^k \cup\{E_d\}.$$

2. Модели потоков стираний

По аналогии с известным понятием потока ошибок введем понятие потока стираний [6]. Потоком стираний назовем последовательность $\{e_i\}$ элементов из алфавита {0, *}, длина которой совпадает с длиной закодированных данных, передаваемых по каналу. Символ 0 потока означает отсутствие стирания, а символ * – его наличие в соответствующей позиции передаваемых данных.

Рассмотрим модель генерации потока независимых стираний. Входными параметрами модели являются: вероятность p_* появления стираний, длина L потока. На выходе модели формируется поток стираний $\{e_i\}$, $i=\overline{1,L}$, элементы которого генерируются равновероятно и независимо. Символ стирания * появляется в потоке ошибок с вероятностью p_*, нулевой символ появляется с вероятностью $1-p_*$.

Для организации потока группирующихся стираний модифицируем известную модель Гилберта потока группирующихся ошибок [3]. Будем считать, что канал связи может находиться в двух состояниях: в «хорошем», когда стирания невозможны, и в «плохом», когда стирания происходят с некоторой постоянной вероятностью ε. Способ переключения между состояниями канала определяется матрицей переходных вероятностей

$$P=\begin{pmatrix} p_{gg} & p_{gb} \\ p_{bg} & p_{bb} \end{pmatrix}, \tag{3}$$

где p_{xy} – вероятность перехода из состояния x в состояние y. Символ g отвечает за «хорошее» состояние, символ b – за «плохое» состояние. Заметим, что для этих вероятностей справедливы следующие зависимости:

$$p_{gg}+p_{gb}=p_{bb}+p_{bg}=1;$$
$$p_{gg}\gg p_{gb},\ \ p_{bb}\gg p_{bg}.$$

Очевидно, что все оценки, известные для потока ошибок, сгенерированного с помощью модели Гил-

берта, справедливы и для случая потока стираний. Далее используем среднюю вероятность ошибочного приема элемента, которая согласно [3] вычисляется следующим образом:

$$q = \varepsilon * \frac{p_{gb}}{p_{bg} + p_{gb}}. \quad (4)$$

Таким образом, входными параметрами модели группирующихся стираний являются: матрица переходных вероятностей (3), вероятность стирания в плохом состоянии канала ε и длина потока L. На выходе модели строится поток $\{e_i\}$, $i = \overline{1, L}$ группирующихся стираний, вероятность символа стирания в котором оценивается формулой (4).

3. Метод равновесных столбцов и его модификация

В [1] предложен метод кодирования двоичных данных, позволяющий исправлять стирания и названный методом равновесных столбцов. Метод описывает способ построения кодирующих матриц $[n, k]_2$-кода, состоящих из n случайных векторов-столбцов, длины k и веса $w = (k+1)/2$. Особенностью матрицы, построенной методом РС, является тот факт, что любые k столбцов этой матрицы с большой вероятностью являются линейно-независимыми.

Рассмотрим кодирующую $(k \times n)$-матрицу G. Будем полагать, что $rank(G) = k$. Обозначим через D_i множество всех подматриц, полученных из G путем выбора i столбцов (такие подматрицы будем называть столбцовыми подматрицами). Через R_i, $i = k, \ldots, n$, обозначим множество подматриц, полученных аналогичным способом, которые имеют ранг k. Заметим, что любому элементу множества R_k соответствует некоторая информационная совокупность, а любой элемент множеств R_i, $i = k+1, \ldots, n$, включает в себя столбцовую подматрицу размера $k \times k$, соответствующую информационной совокупности. Нетрудно видеть, что отношение $\rho_i = |R_i|/|D_i|$ соответствует вероятности успешного декодирования [14] информационного сообщения a из i полученных без стираний координат вектора b.

Вектор

$$\overline{f} = (f_k, f_{k+1}, \ldots, f_n), f_i = |R_i| / |D_i|, \ i = \overline{k, n} \quad (5)$$

называется вектором вероятностей успешного декодирования для кодирующей $(k \times n)$-матрицы G [1, 15]. Вектор вероятностей успешного декодирования (ВВУД) используется как инструмент априорной оценки результата декодирования для кодирующей матрицы G. Другими словами, ВВУД можно рассматривать как критерий качества кодирующей матрицы, а значения f_i, $i = \overline{k, n}$, – как вероятность успешного декодирования зашумленного кодового вектора $\overline{c}' = \overline{c} + \overline{e}$, полученного из линии связи и содержащего i нестертых элементов.

Приведем пример кодирующей матрицы $[10, 5]_2$-кода, построенной согласно методу равновесных столбцов:

$$G = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 & 1 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 & 1 & 1 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 1 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 & 0 & 1 & 1 & 1 \end{pmatrix}.$$

Для приведенной матрицы ВВУД выглядит следующим образом:

$$\overline{f_{10,5}} = (0,643; 0,929; 1; 1; 1; 1).$$

Кодирование состоит в умножении информационного вектора $\overline{m} \left(\in F_q^k\right)$ на кодирующую матрицу G:

$$\overline{c} = \overline{m} G,$$

где $\overline{c} \ \left(\in F_q^n\right)$ – кодовый вектор длины n.

Процесс декодирования состоит в применении метода информационных совокупностей [1, 14], адаптированного для каналов со стираниями. Отметим, что процесс декодирования является вероятностным: чем больше стираний повредило кодовое слово, тем меньше вероятность восстановления информационного слова (подробнее в [1]).

В модифицированном методе равновесных столбцов предполагается, что в качестве левой части кодирующей матрицы используется единичная подматрица, а другие столбцы матрицы заполняются согласно методу равновесных столбцов. Использование единичной подматрицы помогает упростить процесс декодирования. Такой прием встречается также в работах [7, 8], посвященных другим методам борьбы со стираниями. Если стирания не попали на те координаты кодовых сообщений, которые соответствуют столбцам единичной матрицы, то процесс декодирования можно выполнить тривиальным образом – просто скопировав нужные координаты кодового вектора в искомый информационный вектор. Приведем пример кодирующей матрицы $[10, 5]_2$-кода, построенной согласно модифицированному методу равновесных столбцов:

$$G = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 1 \end{pmatrix}.$$

Для приведенной матрицы ВВУД выглядит следующим образом: $\overline{f_{10,5}} = (0,643; 0,929; 1; 1; 1; 1)$.

Для метода равновесных столбцов представляется полезным получить теоретическую оценку вероятности P_e неверного декодирования данных, принятых из

стирающего канала. Рассмотрим известную теоретическую формулу оценки вероятности ошибки декодирования P_e данных $[n, k, d]_2$-кодом, исправляющим t ошибок в кодовом слове, при передаче закодированных данных по двоичному симметричному каналу с вероятностью ошибки p [17]:

$$P_e = 1 - \sum_{i=0}^{t} C_n^i p^i \left(1-p\right)^{n-i}, \quad (6)$$

здесь значение C_n^i соответствует количеству векторов ошибок длины n и веса i, которые код может исправить, а произведение $p^i(1-p)^{n-i}$ оценивает вероятность появления i ошибочных и $(n-i)$ верных бит в полученном из канала кодовом слове. На основе (6) в работе эмпирическим путем получена формула теоретической оценки вероятности P_e неверного декодирования методом равновесных столбцов данных, полученных из стирающего канала, использующая вероятности из (5):

$$P_e = 1 - \sum_{i=0}^{t} f_{n-i} C_n^i p^i \left(1-p\right)^{n-i}, \quad (7)$$

где p – вероятность стирания в канале, t – максимальное количество стираний, которое может исправить код (см. (1)), f_{n-i} – коэффициент вектора $\overline{f}$ вероятности успешного декодирования, содержащий вероятность успешного декодирования при получении i нестертых компонент кодового вектора.

4. Имитационная модель

В работе создано программное средство, реализующее имитационную модель цифрового канала с защитой от стираний [16]. Построенная модель предназначена для проведения экспериментов по исследованию качества передачи данных по стирающему каналу. Программа позволяет пользователю установить входные параметры эксперимента: тип и параметры потока стираний, метод защиты от стираний и его параметры, количество экспериментов. В ходе проведения экспериментов имитационная модель генерирует случайным образом последовательность информационных сообщений, кодирует их, генерирует поток стираний и аддитивно добавляет стирания к закодированным сообщениям (см. (2)), затем данные со стертыми элементами в модели декодируются. Далее восстановленные сообщения сравниваются с соответствующими исходными информационными сообщениями. Если результат и оригинал совпадают, то передача сообщения считается успешной, в противном случае – неуспешной. Фиксируется количество успешных и неуспешных результатов для каждого эксперимента.

Рассмотрим использованные в программной реализации алгоритмы генерации потоков стираний.

Алгоритм 4.1. Генерация потока равномерных стираний.

Вход: вероятность стирания в канале p_*, длина кодового слова n, число кодовых слов s.

Выход: поток стираний $e=(e_1, e_2, \ldots, e_{n*s})$, средняя вероятность q стирания в потоке ошибок.

1. Цикл: $i=1,\ldots,n*s$.

Для каждого элемента e_i генерируется случайное число $\omega \in [0, 1]$. Если $\omega \le p_*$, то $e_i = *$, иначе $e_i = 0$.

2. Вычисление вероятности стирания q в сгенерированном потоке ошибок e.

Конец алгоритма.

Алгоритм 4.2. Генерация потока группирующихся стираний.

Вход: параметры модифицированной модели Гилберта ε, p_{gb}, p_{bg}, длина кодового слова n, число кодовых слов s.

Выход: поток стираний $e=(e_1, e_2, \ldots, e_{n*s})$, средняя вероятность q стирания в потоке ошибок.

1. Текущее состояние канала – «хорошее».

2. Цикл: $i=1,\ldots,n*s$, для каждого элемента e_i выполняется:

2.1. Определение нового текущего состояния: генерируется случайное число $\omega \in [0, 1]$, если ω не превосходит вероятность перехода из текущего состояния в противоположное, то считается, что канал остается в текущем состоянии, иначе переходит в противоположное.

2.2. Если состояние канала «плохое», то генерируется случайное число $\omega \in [0, 1]$, если $\omega \le \varepsilon$, то $e_i = *$, иначе $e_i = 0$.

2.3. Если состояние модели «хорошее», то $e_i = 0$.

3. Вычисление вероятности стирания q в сгенерированном потоке ошибок e.

Конец алгоритма.

Отметим, что при использовании модели независимых стираний построить последовательность с заданной вероятностью стирания p_* довольно просто. А именно, если генерируемая последовательность довольно длинная, то итоговая вероятность стирания q в ней равна входному параметру модели p_*. В случае модели группирующихся ошибок создание потока ошибок с заданной вероятностью стирания p_* происходит сложнее по следующим причинам. Во-первых, формула (4) оценки итоговой вероятности стирания по параметрам модели является приблизительной; во-вторых, по заданному значению p_* остальные параметры модели ε, p_{gb}, p_{bg} вычисляются неоднозначно. В связи с этим при проведении экспериментов параметры модели генерировались случайно, но так, чтобы выполнялись условия (3.4). Далее по уже построенному потоку ошибок вычисляется итоговая вероятность стирания q.

5. Экспериментальное исследование

С использованием созданной имитационной модели проведены эксперименты по исследованию корректирующих способностей метода равновесных столбцов и его модификации по отношению к стира-

ниям разного типа. Рассмотрим алгоритм проведения эксперимента.

Вход: параметры кода n, k, количество экспериментов *length*.

Выход: набор следующих значений:

- параметры моделей потоков стираний;
- доля успешно декодированных кодовых слов для группирующихся стираний;
- доля успешно декодированных кодовых слов для равномерных стираний;
- вероятность стирания в потоке (равна для обоих методов генерации потока стираний).

1. Для заданного количества экспериментов *length* генерируем случайные параметры одного эксперимента: ε, p_{gb}, p_{bg}.

2. Цикл по количеству экспериментов:

2.1. Запуск итерации для случая группирующихся стираний и получение итоговой вероятности q стираний.

2.2. Запуск итерации для случая независимых стираний с вероятностью q.

2.3. Сохранение результатов эксперимента.

Конец алгоритма.

На каждой итерации алгоритма создается канальный кодек с параметрами n, k и проводится 1000 экспериментов, в каждом из которых выполняются следующие действия:

1. Генерируется информационное сообщение $\overline{m}$ длины k.

2. Вектор $\overline{m}$ кодируется заданным методом, получаем кодовое сообщение $\overline{c}$.

3. Создается поток стираний $\overline{e}$.

4. Стирания накладываются на сообщение $\overline{c}' = \overline{c} + \overline{e}$.

5. Результат декодирования вектора $\overline{c}'$ записывается в $\overline{m}'$.

6. Если $\overline{m}' = \overline{m}$, то увеличивается счетчик успешных попыток декодирования.

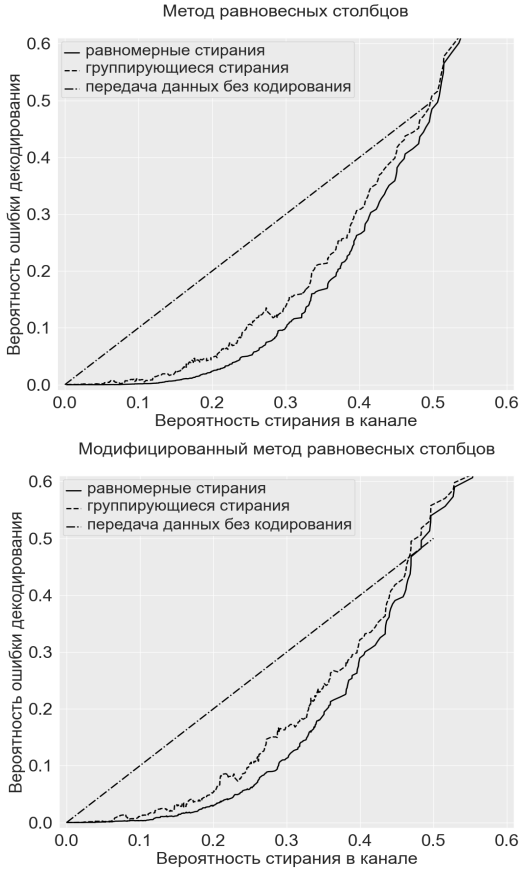
7. Вычисляется частота успешного декодирования.

Конец итерации.

5.1. Сравнение корректирующих способностей методов РС и МРС

На рис. 1*а*, *б* представлены результаты экспериментов, проведенных с использованием метода равновесных столбцов и его модификации (параграф 3), параметры использованных при этом кодов $[10,5]_2$. По оси абсцисс показана вероятность стирания в канале, по оси ординат – вероятность ошибки декодирования. Пунктирная кривая на графике соответствует результатам, полученным для случая равномерных стираний, а сплошная кривая – для случая группирующихся стираний.

Очевидно, что с увеличением вероятности стирания в канале увеличивается вероятность неуспешного декодирования. По всем проведенным экспериментам при группирующихся стираниях вероятность ошибки декодирования больше либо равна вероятности ошибки в случае независимо распределённых стираний. Кривые на графиках пересекают прямую, соответствующую передаче данных без кодирования, в случае метода РС при вероятности стираний $\approx 0{,}5$ в канале, а в случае метода МРС чуть раньше – при вероятности $\approx 0{,}47$. Т.е. после достижения этих порогов методы кодирования становятся неэффективными. Заметим, что если бы эти графики были связаны с двоичными помехоустойчивыми кодами, которые исправляют ошибки, то при вероятности ошибки в канале, превышающей значение 0,5, можно было бы инвертировать зашумленные данные и улучшить результат декодирования. Однако в случае со стираниями такой прием, очевидно, не может быть использован.



а)

б)

Рис. 1. Результаты экспериментального исследования для $[10,5]_2$-кода для методов РС и МРС

5.2. Сравнение качества декодирования методов в случае группирующихся и независимых стираний

Сравним между собой способности исправлять стирания двух рассматриваемых методов кодирования в случае $[10,5]_2$-кода. Из графиков, представленных на рис. 2*а*, видно, что метод РС эффективнее справляется с декодированием независимых стираний (ему соответствует пунктирная кривая). Оба метода практически одинаково справляются со случаем группирующихся стираний (рис. 2*б*), линии графиков

переплетаются, нельзя выделить явного лидера. Напомним, что в среднем скорость работы декодера модифицированного метода выше, чем у метода РС. Кодирующая матрица метода МРС содержит единичную подматрицу. Если стирания не попали на координаты, соответствующие столбцам единичной матрицы, то декодирование состоит в копировании информационного сообщения из начальных бит кодового слова.

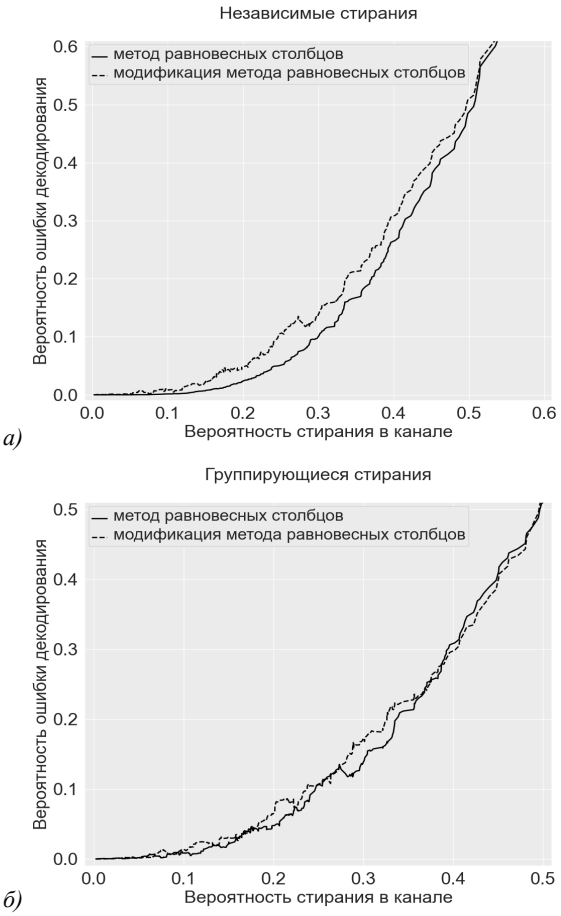


Рис. 2. Сравнение модифицированного и обычного метода равновесных столбцов с параметрами (5, 10) в случае простых и группирующихся стираний

На рис. 3 показаны результаты экспериментов для $[12,6]_2$ (рис. 3*а*) и $[14,7]_2$-кодов (рис. 3*б*). Полученное взаимное расположение кривых на графиках, связанных с методом РС и методом МРС, аналогично результатам, полученным для $[10,5]_2$-кода. Однако в целом результаты декодирования хуже, чем в случае $[10,5]_2$-кода, т.к. кривые быстрее уходят наверх.

5.3. Исследование качества теоретической оценки ошибки декодирования

Теоретическая оценка (7) построена для случая равномерных независимых стираний. На рис. 4 видно, что теоретические и экспериментальные результаты, оценивающие вероятность ошибки декодирования, для случая $[10,5]_2$-кода хорошо согласуются. Максимальное отклонение составляет 0,031. Среднеквадратичное отклонение – 4,54e-05. Для кодов с другими параметрами получены аналогичные результаты, что позволяет сделать вывод о том, что теоретическая оценка довольно точна.

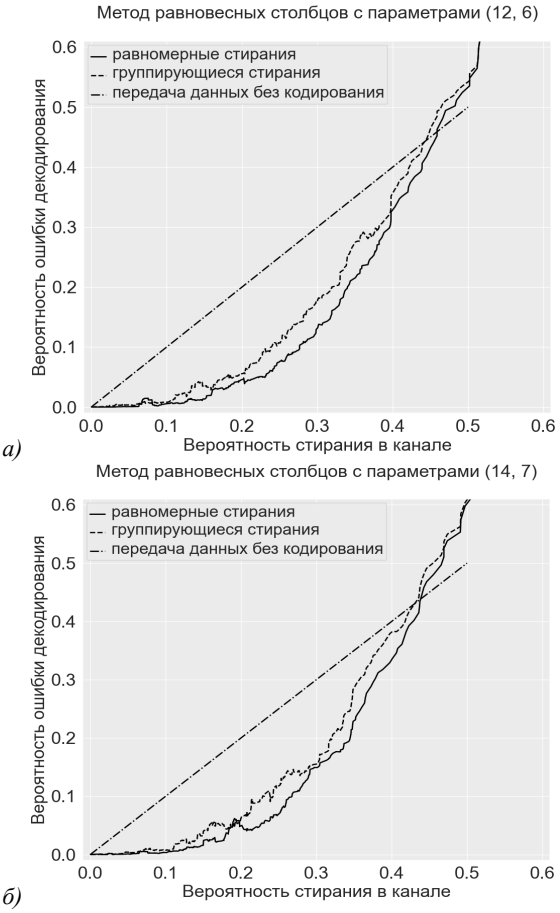


Рис. 3. Сравнение типов моделей потоков стираний для методов равновесных столбцов с параметрами (12,6) и (14,7)

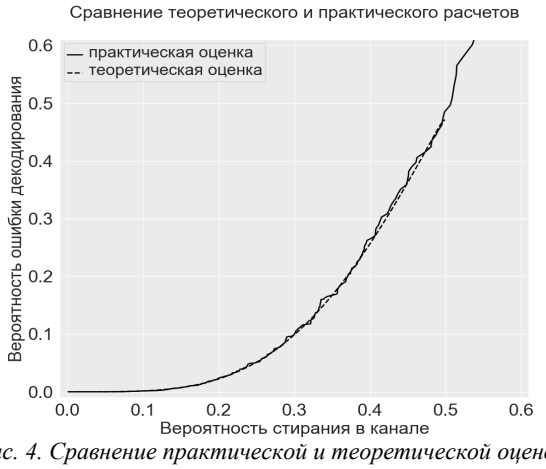


Рис. 4. Сравнение практической и теоретической оценок качества декодирования для метода равновесных столбцов для $[10,5]_2$-кода в случае равномерных стираний

5.4. Подбор параметров кода для случая группирующихся стираний

При заданной вероятности p_* группирующихся стираний в канале и желаемой вероятности ошибки

декодирования p_{dec} теоретически подобрать параметры кода для метода равновесных столбцов можно в два этапа. На первом этапе выбрать такие параметры кода n, k и ВВУД f, чтобы выполнялось равенство (7), где $p=p_*$, $P_e=p_{dec}$. Затем, на втором этапе, увеличить избыточность метода равновесных столбцов, добавив один столбец к порождающей матрице, т.е. перейти к использованию $[n+1,k]_2$-кода. Добавленная избыточность компенсирует ухудшение качества декодирования, вызываемое группирующимися ошибками. На рис. 5 приведены графики, оценивающие ошибки декодирования и построенные для $[9,5]_2$ -кода и равномерных ошибок, а также для $[10,5]_2$-кода и группирующихся ошибок.

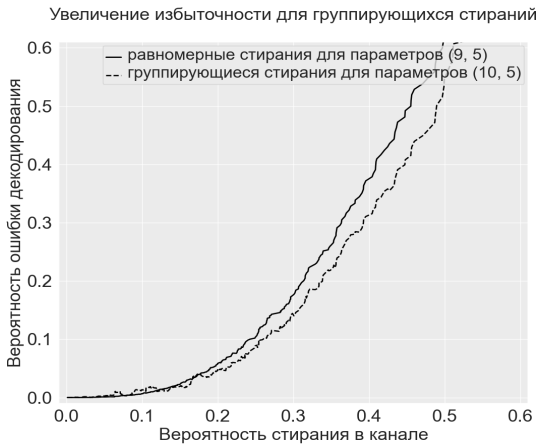


Рис. 5. Пример увеличения избыточности для группирующихся стираний на 1 столбец

5.5. Выводы по результатам проведенных экспериментов

1. Тип потока стираний значительно влияет на качество восстановления данных. А именно, при одной и той же вероятности стирания в канале в случае группирующихся стираний вероятность ошибки декодирования выше, чем в случае независимых равномерно распределенных стираний.

2. В работе не указаны примеры каналов передачи, для которых модифицированная модель Гильберта потока стираний является адекватной, т.е. корректно отображает свойства канала с точки зрения появления стираний. Но проведенные эксперименты ясно показывают, что группирование стираний влияет на корректирующую способность методов защиты, следовательно, группирование стираний следует учитывать, и, возможно, следует разрабатывать новые модели потоков стираний.

3. Предложенная теоретическая оценка вероятности ошибки декодирования для методов РС близка к вероятности, полученной в результате экспериментов в случае независимых ошибок.

4. Для борьбы с группирующимися стираниями эффективно закладывать дополнительную избыточность в код, подобранный с помощью теоретической оценки (7).

5. Метод равновесных столбцов демонстрирует падение качества декодирования при увеличении параметров кода. Это связано со сложностью построения «хороших» порождающих матриц. Кроме этого, при увеличении кода усложняется алгоритм декодирования, в основном из-за увеличения времени на процедуру обращения матриц, т.к. сложность алгоритма обращения кубическая.

6. Отметим, что коды, использованные в примерах, имеют довольно небольшую длину. В настоящее время в реальных системах связи используются коды, длина которых может составлять несколько тысяч бит. Однако в методах РС и МРС лучшую корректирующую способность показывают короткие коды. При необходимости можно увеличить длину кодовых слов с помощью механизма каскадирования кодов [18]. А применение перемежителей в каскадах позволит снизить влияние пакетов стираний на передаваемые кодовые слова, что повысит вероятность корректного декодирования. Заметим, что каскадирование не увеличивает сложность декодирования, но увеличивает задержку между отправкой и получением сообщения. Очевидно, что нельзя сформулировать общую рекомендацию по количеству кодов и перемежителей в каскаде, их взаимному расположению и параметрам. Подбор такого каскада к конкретному каналу требует проведения имитационных экспериментов.

Заключение

В работе создана модель группирующихся стираний, построено программное средство, имитирующее работу стирающего канала передачи данных. На основе имитационной модели проведены эксперименты, показавшие зависимость способности по исправлению стираний метода равновесных столбцов и его модификации от интенсивности и типа стираний в канале передачи, а также от параметров методов кодирования. Построена оценка вероятности ошибки декодирования, учитывающая вероятность стираний в канале передачи и параметры метода кодирования РС. Полагаем, что результаты исследования могут быть актуальны для разработчиков каналов и сетей связи.

References

[1] Aydarkin EE, Deundyak VM. Construction of coding matrices with equilibrium columns for using in channels with deletion [In Russian]. Telecommunications 2020; 3: 11-17.

[2] Gabidulin EM, Pilipchuk NI, Bossert M. Decoding of random network codes. Probl Inf Transm 2010; 46(4): 300-320. DOI: 10.1134/S0032946010040034.

[3] Gilbert EN. Channel throughput with error packets [In Russian]. Kiberneticheskii Sbornik 1964; 9: 109-122.

[4] Gligoroski D, Kralevska K. Families of optimal binary non-MDS erasure codes. 2014 IEEE Int Symposium on Information Theory 2014: 3150-3154. DOI: 10.1109/ISIT.2014.6875415.

[5] Koetter R, Kschischang FR. Coding for errors and erasures in random network coding. IEEE Trans Inf Theory 2008; IT-54(8): 3579-3591.
[6] Deundyak VM, Mayevskiy AE, Mogilevskaya NS. Methods of error-correcting data protection [In Russian]. Rostov-on-Don: SFEDU Publishing; 2014.
[7] Al-Shaikhi A, Ilow J. Design of packet-based block codes with shift operators. EURASIP J Wirel Commun Netw 2010; 2010: 263210. DOI: 10.1155/2010/263210.
[8] Pan VY. Matrix structure and loss-resilient encoding/decoding. Comput Math with Appl 2003; 46: 493-499. DOI: 10.1016/S0898-1221(03)90041-1.
[9] Silva D, Kschischang FR, Koetter R. A rank-metric approach to error control in random network coding. IEEE Trans Inf Theory 2008; IT-54(9): 3951-3967. DOI: 10.1109/TIT.2008.928291.
[10] Aydarkin EE, Deundyak VM. Channel-network cascade for packet and symbol erasures in binary linear network. J Comp Eng Math 2020; 7(2): 3-14. DOI: 10.14529/jcem200201.
[11] Valiska J, Hrušovský B, Marchevsky S, Pillár S. Error models simulations in transmission channels using network simulator environment. Acta Electrotechnica et Informatica 2012; 12(2): 51-58. DOI: 10.2478/v10198-012-0019-1.
[12] Maltsev GN, Dzhumkov VV. A generalized model of a discrete communication channel with grouping errors [In Russian]. Information and Control Systems 2013; 1: 27-33.
[13] Kolesnik VD. Coding in the transmission and storage of information (Algebraic theory of block codes) [In Russian]. Moscow: "Vysshaya Shkola" Publisher; 2009.
[14] Evseev GS. On the complexity of decoding linear codes [In Russian]. Probl Inf Transm 1983; 19(1): 3-8.
[15] Trullos-Cruces O. Exact decoding probability under random linear network coding. IEEE Commun Lett 2011; 15(1): 67-69. DOI: 10.1109/LCOMM.2010.110310.101480.
[16] Aydarkin EE, Mogilevskaya NS Program for modeling data transmission in channels with anti-erasure protection based on the equal-weight columns method [In Russian]. Certificate of State Registration of the Computer Program No. 2021611988 of March 2, 2021.
[17] Morelos-Zaragoza RH. The art of error correcting coding. 2nd ed. Hoboken: John Wiley and Sons Inc; 2006. ISBN: 978-0-470-01558-2.
[18] Barinov AY. Movement in channel coding: properties, structure, specifics applications. J Radio Electron 2019; 1. Source: ⟨http://jre.cplire.ru/jre/jan19/13/text.pdf⟩. DOI: 10.30898/1684-1719.2019.1.13.

Сведения об авторах

Айдаркин Евгений Евгеньевич, 1996 года рождения, в 2017 году окончил Южный федеральный университет по специальности 09.06.01 «Прикладная математика и информатика», аспирант, работает Deep Learning Engineer в eToolkit Inc, программистом в ЮФУ. Область научных интересов: помехоустойчивое кодирование, программирование, машинное обучение, глубокое обучение, компьютерное зрение. E-mail: *aydarkin@sfedu.ru* .

Могилевская Надежда Сергеевна, 1979 года рождения, окончила Донской государственный технический университет, в настоящее время доцент Института математики, механики и компьютерных наук им. И.И. Воровича ЮФУ. Область научных интересов: помехоустойчивое кодирование, распределенные хранилища информации. E-mail: *nmogilevskaya@sfedu.ru* .

ГРНТИ: 20.53.23
Поступила в редакцию 11 марта 2022 г. Окончательный вариант – 13 мая 2022 г.

Experimental study of a matrix method of equal-weight columns correcting ability to protect data from erasure

E.E. Aydarkin [1], N.S. Mogilevskaya [1]
[1] Southern Federal University, I.I. Vorovich Institute of Mathematics, Mechanics and Computer Science, Rostov-on-Don

Abstract

The paper investigates the ability of the equal-weight columns method to resist grouping erasures. The Gilbert model for error generation flow is adapted for the case of erasures. A simulation model of the erasure-correction channel is constructed with the possibility of choosing the type of erasure and the method of protection against erasure. With the help of this model, an experimental study of the equal-weight column method and its modifications is conducted and a detailed analysis of the results with conclusions for developers of networks and data transmission channels is carried out. An estimate of the decoding probability is constructed. A method of dealing with clustering erasures by using additional redundancy is proposed.

Keywords: erasure, error-correction data transmission channel, grouped erasures, Gilbert model, equal-weight columns method.

Citation: Aydarkin EE, Mogilevskaya NS. Experimental study of a matrix method of equal-weight columns correcting ability to protect data from erasure. Computer Optics 2022; 46(5): 840-847. DOI: 10.18287/2412-6179-CO-1122.

Authors' information

Evgeny Evgenievich Aidarkin, born in 1996, graduated from Southern Federal University in 2017, specialty 09.06.01 "Applied Mathematics and Computer Science", postgraduate student, works Deep Learning Engineer at eToolkit Inc. Research interests: error-correcting coding, programming, machine learning, deep learning, computer vision.
E-mail: *aydarkin@sfedu.ru* .

Nadezhda Sergeevna Mogilevskaya, born in 1979, graduated from Don State Technical University, currently an associate professor at the I.I. Vorovich Institute of Mathematics, Mechanics and Computer Sciences of the Southern Federal University. Research interests: error-correcting coding, distributed information storage.
E-mail: *nmogilevskaya@sfedu.ru* .

Received March 11, 2022. The final version – May 13, 2022.