

Практическое сравнение некоторых декодеров для кодов, устойчивых к стираниям: скорость против корректирующей способности

Е.Е. Айдаркин, Н.С. Могилевская

Южный федеральный университет, Ростов-на-Дону

Аннотация: Работа посвящена поиску эффективного метода декодирования для нового класса бинарных кодов, корректирующих стирания. Рассматриваемые коды задаются кодирующей матрицей, имеющей ограничения на веса столбцов (МРСт-коды). Для работы с построенными кодами используются декодер по информационным совокупностям и декодер по принципу распространения доверия, адаптированные для случая работы со стираниями. В работе проведены эксперименты по определению скорости декодирования и корректирующей способности этих методов применительно к названным классам помехоустойчивых кодов. В случае МРСт-кодов декодер по принципу распространения доверия значительно выигрывает по скорости по сравнению с декодером по информационным совокупностям, однако незначительно проигрывает по уровню корректирующей способности.

Ключевые слова: каналы со стираниями, распределенные отказоустойчивые системы хранения данных, код с равновесными столбцами, декодер по информационным совокупностям, декодер по принципу распространения доверия, РСт-код, МРСт-код.

Введение. Одно из актуальных направлений теории помехоустойчивого кодирования связано с кодами, корректирующими стирания. Такие коды позволяют восстанавливать частичную потерю данных. Коды, корректирующие стирания, могут быть использованы, например, в распределенных системах хранения данных, где они позволяют восстанавливать данные, утерянные в случае непреднамеренных аппаратных или программных ошибок, что, в свою очередь, повышает надежность и эффективность систем хранения [1 – 3]. Отметим, что в отличие от задачи защиты данных от искажений в линиях передачи, где используются коды с длиной кодового слова в несколько сотен или тысяч элементов, в системах хранения применяются короткие коды с длиной менее 20 элементов (см., например, [4]).

В работе [5] описан быстрый способ построения кодирующих матриц двоичных кодов борьбы со стираниями. Способ позволяет выбирать длину и размерность кода независимо друг от друга. Описанный способ построения

матриц эффективен для генерации кодов с небольшой длиной и размерностью. Все столбцы произвольной кодирующей матрицы, построенной этим способом, имеют фиксированный вес. В работе [5] они названы равновесными столбцами, а коды, задаваемые такими кодирующими матрицами – равновесными кодами (РСт коды). В [5] также построена модификация РСт кодов, кодирующая матрица этих кодов, состоит из двух частей: единичная подматрица и подматрица с равновесными столбцами. Эта модификация РСт кодов названа МРСт кодами.

В работе [6] приведены результаты исследования корректирующей способности кодов РСт и МРСт по отношению к группирующимся и независимым стираниям различной интенсивности. Для декодирования в [6] использован способ декодирования по информационным совокупностям, адаптированный для случая стираний. Этот способ декодирования показал ожидаемую корректирующую способность, однако, для практических приложений хотелось бы найти способ декодирования с меньшей вычислительной сложностью.

Напомним, что в теории помехоустойчивого кодирования алгоритмы декодирования разделяют специальные и универсальные [7]. Специальные декодеры разрабатываются для узкого класса помехоустойчивых кодов. Они используют структурные свойства кода для упрощения процесса декодирования. Обычно специальные декодеры вычислительно менее сложны по сравнению с универсальными декодерами. Универсальные декодеры работают с произвольными линейными кодами и обладают либо высокой вычислительной сложностью, либо высокими требованиями к объему используемой памяти [8].

Есть распространённое мнение, что универсальные декодеры применимы только в учебных целях и для взлома кодовых криптосистем. Однако, во-первых, по сей день разрабатываются модификации известных

универсальных методов декодирования, направленные на улучшение свойств декодеров и расширение сферы их применения (см., например, для декодера по информационным совокупностям работы [9 – 11], для декодера по принципу распространения доверия [12 – 14]), а, во-вторых, для коротких кодов вычислительная сложность универсальных декодеров может быть приемлемой.

Произвольная кодирующая матрица МРСт-кодов не обладает какой-либо особой алгебраической структурой, следовательно построение специальных эффективных по сложности или требованиям к памяти декодеров для указанных кодов с произвольными параметрами затруднительно.

Цель данной работы состоит в адаптации декодера по принципу распространения доверия для стирающего канала передачи данных и проведении сравнительного анализа скорости работы и корректирующей способности декодера по информационным совокупностям и декодера по принципу распространения доверия для МРСт кодов.

Необходимые теоретические сведения и обозначения. Обозначим F_q поле Галуа мощностью q , а множество векторов длиной r , составленных из элементов поля F_q обозначим F_q^r .

Рассмотрим помехоустойчивый линейный блочный $[n, k, d]_q$ -код C , где n – его длина, k – размерность ($k < n$), d – минимальное кодовое расстояние [7]. Пусть код C задан кодирующей $k \times n$ матрицей G , состоящей из элементов поля F_q , $\text{rank}(G) = k$. Кодирование определяется формулой

$$\bar{c} = \bar{t}G, \quad (1)$$

где $\bar{c} = (c_1, \dots, c_n) \in C (\subset F_q^n)$ – кодовое слово, а $\bar{t} \in F_q^k$ – информационный вектор.

В результате передачи кодового вектора по стирающему каналу передачи данных среди элементов этого вектора могут появиться стертые

элементы. Обозначим их символом $*$, тогда, можно сказать, что на вход декодера поступают векторы из поля $F_{q^*} = F_q \cup \{*\}$. Далее рассмотрим используемые в работе декодеры.

Декодер по информационным совокупностям. Множество номеров коэффициентов кодового слова $\gamma = \{j_1, j_2, \dots, j_k\}$, $1 \leq j_1 < j_2 < \dots < j_k \leq n$ называется информационной совокупностью (ИС) $[n, k, d]_q$ -кода C , если коэффициенты $c_{j_1}, c_{j_2}, \dots, c_{j_k}$ однозначно определяют кодовое слово $\bar{c} = (c_1, \dots, c_n) \in C$. Известно, что произвольное множество γ является информационной совокупностью тогда и только тогда, когда является невырожденной матрица G_γ , составленная из столбцов матрицы G , номера которых входят в γ [15].

Декодер по информационным совокупностям в случае стирающего канала основан на следующем факте: если в принятом слове $\bar{b} \in F_{q^*}$ имеются нестертые коэффициенты, позиции которых покрывают хотя бы одну информационную совокупность, то $\bar{b}$ может быть корректно декодирован. А именно, пусть $\bar{b}_\gamma$ – это вектор, составленный из коэффициентов вектора $\bar{b}$ с номерами $\gamma = \{j_1, j_2, \dots, j_k\}$, тогда информационный вектор может быть вычислен по формуле: $\bar{t} = \bar{b}_\gamma G_\gamma^{-1}$.

Наиболее трудоемкой операцией декодера по информационным совокупностям является вычисление обратной матрицы G_γ^{-1} , сложность алгоритма обращения матрицы оценивается как $O(k^3)$. Есть три очевидных подхода к вычислению подобных матриц в процессе декодирования потока кодовых слов, искаженных стираниями. Во-первых, такие матрицы можно

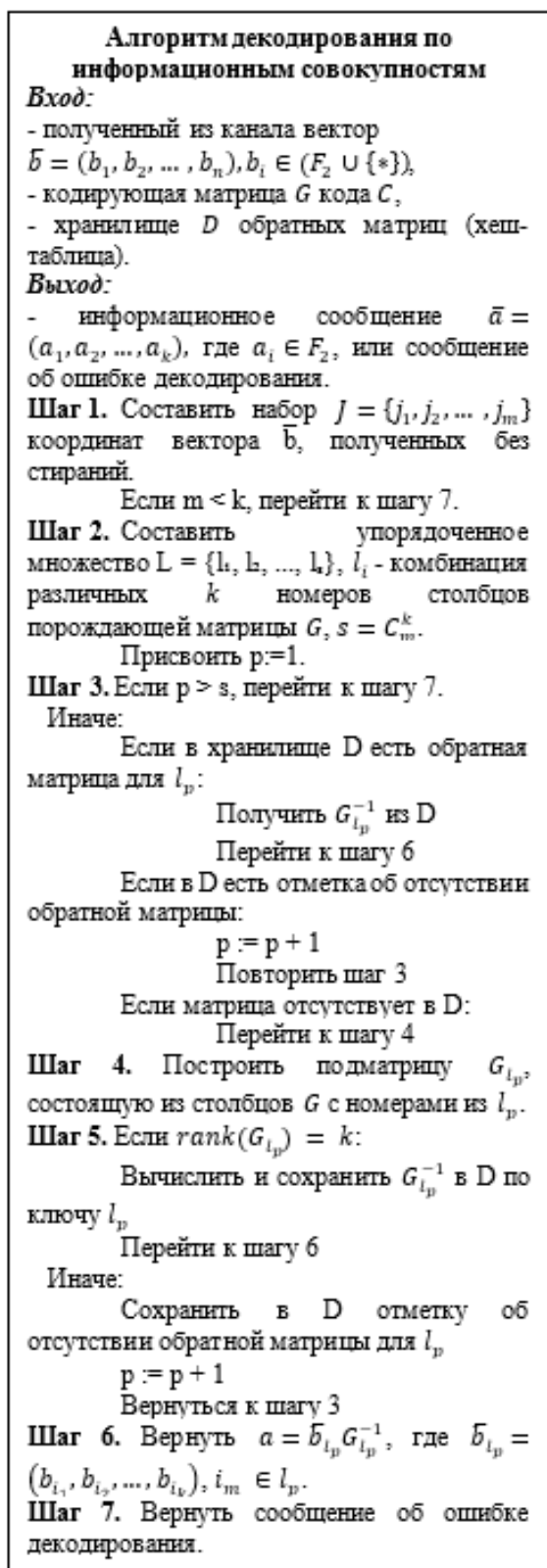


Рисунок 1. Алгоритм декодирования по информационным

вычислять каждый раз, когда в процессе декодирования возникает такая задача. Во-вторых, такие матрицы для всех возможных информационных совокупностей можно вычислить заранее до начала работы декодера, а затем хранить их в памяти компьютера. В-третьих, обратные матрицы, как и информационные совокупности, можно вычислять в процессе декодирования, а затем помещать их в память компьютера. Тогда со временем работы декодера постепенно накопится список всех информационных совокупностей данного кода и соответствующих матриц $G_{l_p}^{-1}$. Далее в работе будем рассматривать третий вариант работы с обратными матрицами декодера по информационным совокупностям.

Алгоритм декодирования (см. рис. 1) в своей работе использует хранилище D , в котором в ходе работы накапливает множество информационных совокупностей кода и соответствующих обратных матриц. Хранилище D – это хэш-таблица ключами которой являются все возможные множества l_p номеров

столбцов кодирующей матрицы G , $|l_p| = k$. Значение хэш-таблицы по ключу l_p зависит от ранга матрицы G_{l_p} , построенной из столбцов кодирующей матрицы G с номерами из ключа l_p : если $\text{rank}(G_{l_p}) = k$, то значением является $G_{l_p}^{-1}$, иначе в значение таблицы записывается специальная метка, указывающая на отсутствие матрицы обратной к G_{l_p} .

Декодер по принципу распространения доверия. Декодеры, работающие по принципу распространения доверия, и их применение к различным кодам, исправляющим ошибки, описаны, например, в [16 – 18]. Алгоритм декодирования, адаптированный для стирающего канала и используемый в работе, показан на рисунке 2. Алгоритм можно интерпретировать как работу с двудольным графом, состоящим из узлов переменных, которые соответствуют коэффициентам полученного кодового слова и узлов проверок, которые соответствуют проверочным уравнениям кода. Алгоритм работает итерационно. На каждой итерации он удаляет один информационный символ из графа и все его связи, после чего значения в других вершинах графа пересчитываются. Алгоритм корректен, так как представляет собой процедуру решения линейных уравнений методом подстановок без этапа преобразования к верхнетреугольному виду [18].

Отметим, что некоторые конфигурации стираний, которые могут быть корректно восстановлены другими декодерами, данным алгоритмом не могут быть исправлены. Однако, этот недостаток компенсируется относительной вычислительной простотой, алгоритм в худшем случае обладает сложностью $O(k^2)$ [20].

Алгоритм работает итеративно, на шагах 3–6 происходит декодирование одного информационного символа. Следовательно, алгоритм завершится за k итераций, или выдаст ошибку декодирования.

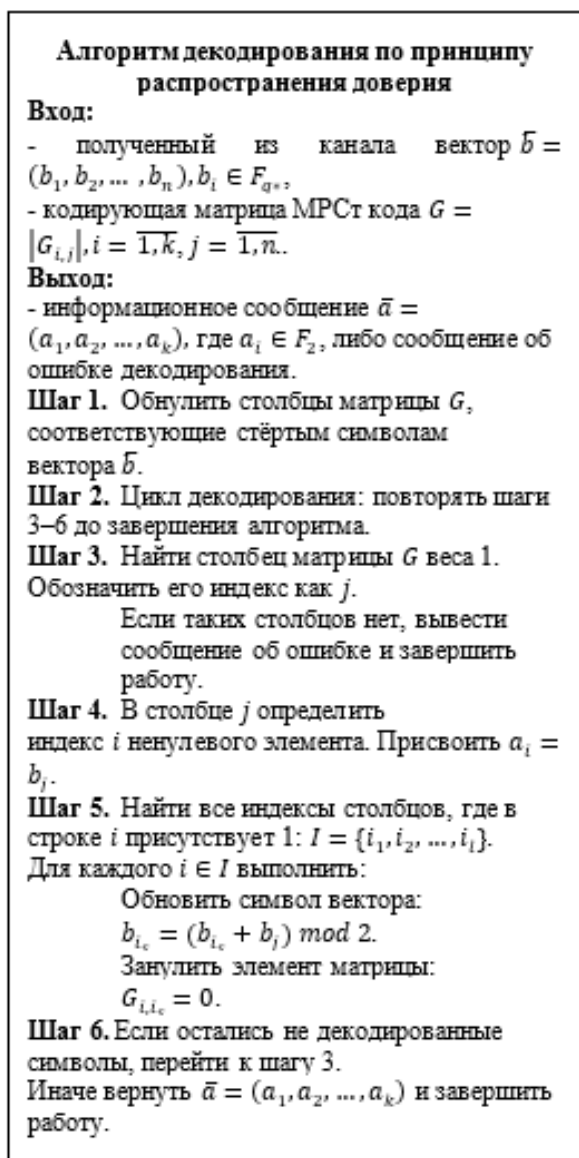


Рисунок 2. Алгоритм декодирования по принципу распространения

Заметим, что обычно при обсуждении декодеров, работающих по принципу распространения доверия, говорят о работе с проверочной матрицей H , но в нашей работе используется кодирующая матрица G кода C . Эти матрицы связаны известным равенством $GH^T = 0$.

Экспериментальное сравнение алгоритмов декодирования.

Описанные выше алгоритмы декодирования программно реализованы и включены в состав имитационной модели стирающего канала передачи данных [5]. Кроме выбора параметров алгебраического кода, алгоритмов его кодирования и декодирования, модель позволяет выбирать параметры потока стираний, а именно, структуру и частоту стираний. Имитационная модель дает

возможность проводить эксперименты с заданными параметрами канала передачи, а также фиксирует параметры и результаты эксперимента.

Рассмотрим схему проведения эксперимента. По заданным параметрам n и k кода генерируется случайная порождающая матрица G . Затем создается s случайных информационных сообщений длины k и s случайных векторов стираний длины n . Стирания генерируются равномерно с заданной вероятностью ρ . Случайные сообщения кодируются (см. (1)). На

последовательность построенных кодовых сообщений аддитивно накладывается последовательность векторов стираний. Каждое из s поврежденных кодовых слов поступает на вход каждого из алгоритмов декодирования. Результат декодирования фиксируется.

В проведенных экспериментах размерность кода k выбиралась последовательно от 5 до 20; длина кода n всегда составляла $2k$, т. е. фиксировалась избыточность кода. Для каждой пары параметров n и k число сообщений $s = 10^4$. Далее приведены результаты некоторых экспериментов. На каждом рисунке представлено два графика, связанных с фиксированными параметрами n и k . Левый график показывает зависимость времени работы декодера от вероятности стирания в канале. Правый график показывает зависимость числа верно декодированных кодовых слов от вероятности стирания в моделируемом канале передачи.

Для краткости при описании результатов параметры n и k используемого кода будем описывать кортежем (n, k) . На рисунке 3 показаны графики, построенные для (10, 5)-кода. Скорость работы декодера по принципу распространения доверия с ростом вероятности стирания изменяется незначительно и плавно. Скорость работы декодера по информационным совокупностям резко растет с увеличением вероятности стирания. При вероятности стирания не превосходящем 0.12 эффективнее по времени работает декодер по принципу информационных совокупностей, затем эффективнее становится второй декодер. Качество декодирования (см. график справа на рис. 1) обоих декодеров практически совпадает с небольшим преимуществом алгоритма 1.

При увеличении длины информационного слова до 8 (рис. 4) наблюдается значительный разрыв в скоростях алгоритмов декодирования. Значительное преимущество наблюдается у алгоритма по принципу распространения доверия. Соответствующая кривая на рисунке 2 находится

близко к 0, но не равна ему (среднее время исполнения 1,11 с). Однако разница в корректирующей способности алгоритмов декодирования становится более значительной по сравнению с меньшей длиной блока. Преимущество здесь на стороне алгоритма, использующего информационные совокупности.

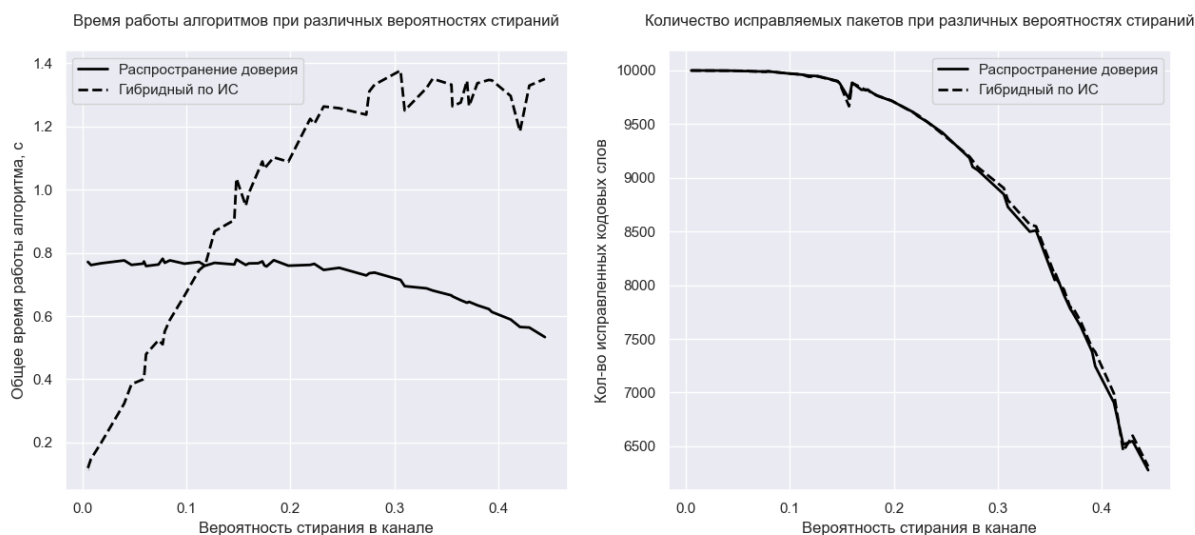


Рисунок 3. Сравнение работы алгоритмов декодирования для (10, 5) МРСТ-кода.



Рисунок 4. Сравнение работы алгоритмов декодирования для (16, 8) МРСТ-кода.

При дальнейшем увеличении длины блока (см. рис. 5) разрыв в скорости работы алгоритмов резко растет, а разница в корректирующей способности кодов сохраняется примерно такой же, как и на рис. 2. Время работы алгоритма декодирования по принципу распространения доверия в среднем равно 1,42 с.

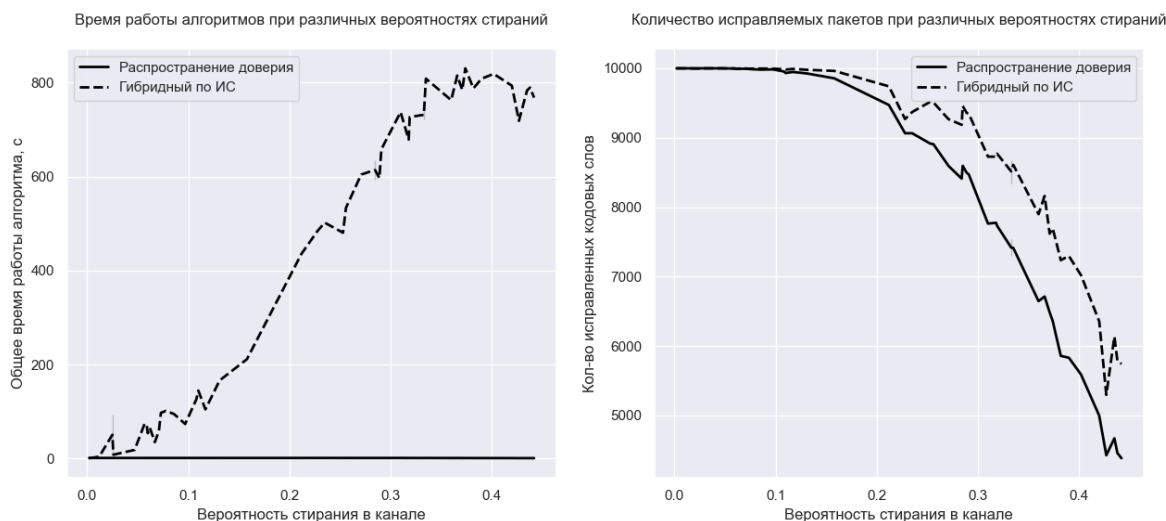


Рисунок 5. Сравнение работы алгоритмов декодирования для (20, 10) МРСТ-кода.

Таким образом, эксперименты показали, алгоритм по принципу распространения доверия более эффективен с точки зрения скорости декодирования, однако его декодирующая способность ниже алгоритмов, основанных на декодировании по информационным совокупностям.

Заметим, что декодер по принципу распространения доверия имеет смысл применять в системах передачи реального времени, с критичными требованиями к задержке и скорости работы системы. А при восстановлении важной информации целесообразнее воспользоваться декодером по информационным совокупностям, обеспечивающим лучшее качество декодирования.

Декодер по информационным совокупностям обладает большей вычислительной сложностью $O(k^3)$ против $O(k^2)$ у метода по принципу

распространения доверия. Однако, к нему легко применить распараллеливание, для одновременной проверки разных информационных совокупностей. Распараллеливание безусловно ускорит работу декодера, но потребует дополнительных вычислительных ресурсов.

Алгоритм 1 во время своей работы постепенно заполняет хеш-таблицу с обратными матрицами. Скорость алгоритма растет с уровнем заполненности хеш-таблицы, а затем фиксируется на некотором постоянном значении. Скорость заполнения хеш-таблицы зависит не только от числа декодированных кодовых слов, но и от вероятности появления стираний в кодовых словах, а также их структуры.

Скорость работы алгоритмов декодирования может отличаться в зависимости от качества кодирующих матриц кода. Отметим, что для кодов МРСт (как и для РСт) качество кодирующих матриц кода определяется с помощью вектора вероятности успешного декодирования [6], а не с помощью более привычного критерия – минимального кодового расстояния.

Результаты исследования, представленные выше, связаны только с МРСт кодами. Отметим, что в начале исследования декодер, работающий по принципу распространения доверия, предполагалась применять, как к МРСт кодам, так и к РСт кодам. К сожалению, в рамках данного исследования адаптировать декодер, работающий по принципу распространения доверия, к РСт кодам не удалось. Из теории известно, что существенное ограничение декодера по принципу распространения доверия состоит в том, что хорошие результаты в коррекции ошибок можно получить только в случае использовании кода с разряженной кодовой/проверочной матрицей. Однако, обычно какой-либо численный критерий по уровню разреженности матрицы, задающей код, не предъявляется. Алгоритм 2 может быть успешно использован в случае МРСт кода, порождающая $(k \times n)$ -матрица которого содержит $\approx k + (n - k)/2$ единичных бит. Алгоритм декодирования по

принципу распространения доверия (см. рис. 2) не может быть успешно использован в случае кода РСт, порождающая $(k \times n)$ -матрица которого содержит приблизительно $kn/2$ единичных бит. Результаты исследования качества работы РСт-кодов с декодером по информационным совокупностям можно найти в [6].

Заключение. В работе представлен сравнительный анализ двух декодеров для МРСт кодов, исправляющих стирания. Оценена вычислительная эффективность (скорость) этих декодеров. Экспериментально исследована корректирующая способность декодеров. Выявлены условия, при которых каждый из декодеров является предпочтительным. Результаты могут быть полезны для разработчиков и исследователей при выборе подходящего метода декодирования для конкретных приложений.

Одним из возможных дальнейших направлений работы, связанной с РСт и МРСт кодами является разработка специализированных декодеров для некоторых кодов этого класса, что может позволит значительно ускорить процесс восстановления данных. За основу специализированного декодера предполагается использовать алгоритм по принципу распространения доверия.

Литература

1. Balaji S.B., Krishnan M.N., Vajha M. Erasure coding for distributed storage: an overview. Science China Information Sciences, 2018, 61. URL: scis.scichina.com/en/2018/100301.pdf
 2. Li A., Yuan X. Design of flexible Compression Transmission Framework for Distributed Storage System Based on Erasure Coding. 5th International Conference on Frontiers Technology of Information and Computer (ICFTIC). Qiangdao. 2023. Pp. 501-504.
-

3. Могилевская Н.С., Юрасов А.И. Анализ симуляторов систем отказоустойчивого хранения. Инженерный вестник Дона. 2025. 2. URL: ivdon.ru/uploads/article/pdf/IVD_49N1y25_mogilevskaya_iurasov.pdf_060aa0bf14.pdf.
 4. Qi Y., Feng D., Hou B. Towards building reliable and cost-efficient distributed storage systems. IEEE Access. 2020. 8. Pp. 157862–157877.
 5. Айдаркин Е. Е., Деундяк В. М. Построение кодирующих матриц с равновесными столбцами для использования в каналах со стираниями. Телекоммуникации. 2020. 3. С. 11-17.
 6. Айдаркин Е. Е., Могилевская Н. С. Экспериментальное исследование корректирующей способности матричного метода равновесных столбцов защиты данных от стираний. Компьютерная оптика. 2022. 46. 5. С. 840-847.
 7. Деундяк В. М., Маевский А. Э., Могилевская Н. С. Методы помехоустойчивой защиты данных. Ростов-на-Дону. Издательство Южного федерального университета. 2014. 309 с.
 8. Айдаркин Е. Е., Могилевская Н. С. О применении декодеров кодов, исправляющих ошибки, в каналах со стираниями. Инженерный вестник Дона. 2023. 12. URL: ivdon.ru/uploads/article/pdf/IVD_99__11_aidarkin_mogilevskaya.pdf_3ec2608a43.pdf.
 9. May A., Ozerov I. On Computing Nearest Neighbors with Applications to Decoding of Binary Linear Codes. Lecture Notes in Computer Science. 2015. 9056. Pp. 203-228.
 10. Bitzer S., Bossert M. On Multibasis Information Set Decoding. IEEE International Symposium on Information Theory. 2022. Pp. 2780-2784.
-

11. Weger V., Battaglioni M., Santini P., Chiaraluce F., Baldi M., Persichetti E. Information set decoding of Lee-metric codes over finite rings. Preprint arXiv. 2020. URL: 10.48550/arXiv.2001.08425.
 12. Sha J., Lin J., Wang Z. Stage-combined belief propagation decoding of polar codes. International Symposium on Circuits and Systems. 2016. Pp. 421-424.
 13. Zhang M.-F., Li Z., Xing L., Liao X. Higher-Order Belief Propagation Correction Decoder for Polar Codes. Entropy. 2022. 24. 534. URL: [scispace.com/pdf/higher-order-belief-propagation-correction-decoder-for-polar-1f2c1eoi.pdf](https://arxiv.org/pdf/higher-order-belief-propagation-correction-decoder-for-polar-1f2c1eoi.pdf).
 14. Nachmani E., Marciano E., Lugosch L., Gross W., Burshtein D., Be'ery Y. Deep Learning Methods for Improved Decoding of Linear Codes. IEEE Journal on Selected Topics in Signal Processing. 2017. 12. Pp. 119-131.
 15. Колесник В. Д. Кодирование при передаче и хранении информации. Москва. Высшая школа. 2009. 549 с.
 16. Shokrollahi A. Raptor codes. IEEE Transactions on Information Theory. 2006. 52. 6. Pp. 2551-2567.
 17. Luby M., Mitzenmacher M., Shokrollahi A., Spielman D. Efficient erasure correcting codes. IEEE Transactions Information Theory. 2001. 47. Pp. 569-584.
 18. Zyablov V. V. and Pinsker M. S. Decoding complexity of low-density codes for transmission in a channel with erasures. Problem Information Transmission. 1974. 10. 1. Pp. 10-21.
 19. Трифонов П. В. Основы помехоустойчивого кодирования. Санкт-Петербург. Университет ИТМО. 2022. 231 с.
 20. Luby M. LT-codes. The 43rd Annual IEEE Symposium on Foundations of Computer Science. Vancouver. BC. 2002. Pp. 271-280.
-

References

1. Balaji S.B., Krishnan M.N., Vajha M. Science China Information Sciences, 2018, 61. URL: scis.scichina.com/en/2018/100301.pdf
 2. Li A., Yuan X. 5th International Conference on Frontiers Technology of Information and Computer (ICFTIC). Qiangdao. 2023. Pp. 501-504.
 3. Mogilevskaya N.S., Yurasov A.I. Inzhenernyy vestnik Dona (Rus). 2025. 2. URL: ivdon.ru/uploads/article/pdf/IVD_49N1y25_mogilevskaya_iurasov.pdf_060aa0bf14.pdf.
 4. Qi Y., Feng D., Hou B. IEEE Access. 8. Pp. 157862–157877.
 5. Aydarkin E. E., Deundyak V. M. Telekommunikatsii (Rus). 2020. 3. Pp. 11-17.
 6. Aydarkin E. E., Mogilevskaya N. S. Komp'yuternaya optika. 2022. 46. 5. Pp. 840-847.
 7. Deundyak V.M., Mayevskiy A.E., Mogilevskaya N.S. Metody pomexoustoychivoy zashchity dannykh [Methods of noise-resistant data protection]. Rostov-na-Donu. Izdatel'stvo Yuzhnogo federal'nogo universiteta. 2014. 309 p.
 8. Aydarkin E. E., Mogilevskaya N. S. Inzhenernyy vestnik Dona. 2023. 12. URL: ivdon.ru/uploads/article/pdf/IVD_99__11_aidarkin_mogilevskaya.pdf_3ec2608a43.pdf.
 9. May A., Ozerov I. Lecture Notes in Computer Science. 2015. 9056. Pp. 203-228.
 10. Bitzer S., Bossert M. IEEE International Symposium on Information Theory. 2022. Pp. 2780-2784.
 11. Weger V., Battaglioni M., Santini P., Chiaraluce F., Baldi M., Persichetti E. Preprint arXiv. 2020. URL: [10.48550/arXiv.2001.08425](https://arxiv.org/abs/2001.08425).
 12. Sha J., Lin J., Wang Z. International Symposium on Circuits and Systems. 2016. Pp. 421-424.
-

13. Zhang M.-F., Li Z., Xing L., Liao X. Entropy. 2022. 24. 534. URL: [scispace.com/pdf/higher-order-belief-propagation-correction-decoder-for-polar-1f2c1eoi.pdf](https://arxiv.org/pdf/higher-order-belief-propagation-correction-decoder-for-polar-1f2c1eoi.pdf).
14. Nachmani E., Marciano E., Lugosch L., Gross W., Burshtein D., Be'ery Y. IEEE Journal on Selected Topics in Signal Processing. 2017. 12. Pp. 119-131.
15. Kolesnik V. D. Kodirovaniye pri peredache i khranении informatsii [Coding during transmission and storage of information]. Moskva. Vysshaya shkola. 2009. 549 p.
16. Shokrollahi A. IEEE Transactions on Information Theory. 2006. 52. 6. Pp. 2551-2567.
17. Luby M., Mitzenmacher M., Shokrollahi A., Spielman D. IEEE Transactions Information Theory. 2001. 47. Pp. 569-584.
18. Zyablov V. V. and Pinsker M. S. Problem Information Transmission. 1974. 10. 1. Pp. 10-21.
19. Trifonov P.V. Osnovy pomekhoustoychivogo kodirovaniya [Fundamentals of noise-resistant coding]. Sankt-Peterburg. Universitet ITMO. 2022. 231 p.
20. Luby M. 43rd Annu. IEEE Symposium on Foundations of Computer Science. Vancouver. BC. 2002. Pp. 271-280.

Дата поступления: 17.08.2025

Дата публикации: 25.09.2025